Application of Time-Scale Techniques to Detection of Epileptiform Activity in the EEG

Hansjerg Gölz, DipPhysics

A thesis presented for the degree of
Doctor of Philosophy
in
Electrical and Electronic Engineering
at the
University of Canterbury,
Christchurch, New Zealand.

November 2000

ABSTRACT

Epilepsy is a neurological disorder for which the electroencephalogram (EEG) is the most important diagnostic tool. Detection of interictal (between-seizure) epileptiform activity in the EEG is, however, complicated by various artifacts and spike-like transient waveforms in the normal background EEG. Approaches to the automatic detection of epileptiform discharges (EDs) require sophisticated signal analysis methods.

We have developed a new system for the detection of EDs – spikes, sharp waves, and spike-and-wave complexes – in the EEG. It is based on the continuous wavelet transform (CWT) and incorporates statistical pattern recognition and 3D spatial source analysis.

Wavelet-based approaches to signal analysis include the discrete wavelet transform (DWT), the CWT, and matching pursuit. Both DWT and CWT are based on a prototype filter, defined by a wavelet function applied at various scales, that can lead to accurate localization in both the time and frequency domains. Wavelet-based methods are more appropriate for the analysis of non-stationary signals – such as the EEG – than the more conventional standard and short-time Fourier transforms. Both the DWT and CWT[1] have been applied previously to the spike detection problem but the CWT with a complex-valued wavelet is considered superior due to translation-invariance and independence from the phase of the transient.

Statistical pattern recognition covers a wide range of classification techniques including artificial neural networks and both linear and quadratic discriminant functions. Linear and quadratic discriminant functions perform well on the separation of two non-Gaussian distributed samples from multidimensional feature spaces. In many medical diagnostic decision problems there is a large number of healthy controls but only a small number of definitively diagnosed patients. The resulting unequal sample sizes pose a problem for classification but can be counter-balanced by a bias term in the discriminant functions. An alternative interpretation of the output of discriminant functions can be gained by using a Bayesian approach to provide an additional measure of confidence in the decision process.

A quasi-regular geodesic sampling grid of hypothetical current sources (dipoles) in the source space (brain volume) leads to a new paradigm for the estimation of source

[1]The CWT has been applied to the spike detection problem with both real- and complex-valued wavelets.

location and source orientation. The orientation estimate forms the basis of a new characteristic spatial feature of epileptogenic spikes.

The wavelet-based system can be separated into two distinct stages. Stage 1 detects candidate epileptiform discharges (CEDs). The magnitude of wavelet coefficients of a single scale of the CWT are evaluated on a logarithmic scale to adaptively estimate the background activity. Sudden deviations from the background trigger CED detections. In Stage 2 features of CEDs are extracted from their multichannel multiscale context. Important features are derived from the multiscale edges and ridges of the CWT. A linear discriminant function with a Bayesian approach is applied to obtain the probability that a single-channel transient is epileptiform. The source orientation of each CED is estimated with the dipole model and another linear discriminant function is used to estimate the spike probability of the multichannel CED. A training set of 53 EEGs (13 epileptiform and 40 normal) was used in the design and training of the system.

The system was evaluated using a test set of 53 EEGs (14 epileptiform and 39 normal). This indicated a mean sensitivity of 57.9%, a mean selectivity of 62.5%, and a mean false detection rate of 6.8 false detections per hour. Detection results compare favourably with those obtained with two other published systems (Hybrid I and Hybrid II) on the same data set. The system is sensitive to weak EDs but fails to reject artifacts in several recordings. It is considered that the system would benefit from the addition of a further stage able to make use of wide-sense temporal context of detections.

In summary, it has been shown that the CWT can function as a very sensitive time-adaptive multichannel detector for CEDs. The multichannel multiscale context of EDs provides important features for discrimination of EDs from artifacts via a linear classifier. When used in conjunction with a source-orientation estimate from a dipole model, the resultant system shows considerable potential as an accurate spike detector for applications in clinical neurophysiology.

ACKNOWLEDGEMENTS

I am grateful to many people for the advice, support, assistance and encouragement they gave me during the course of my research and the preparation of this thesis.

I would like to thank my supervisors, Dr Philip Bones and Dr Richard Jones for their guidance, feedback and encouragement.

I gratefully acknowledge receipt of a University of Canterbury Postgraduate Scholarship.

Finally, I wish to thank my family, friends and flatmates for their support and encouragement.

CONTENTS

PREFACE

The following thesis presents the work done in more than three years as a student in the Department of Electrical and Electronic Engineering. Most of the research and development was conducted in the Department of Medical Physics and Bioengineering at Christchurch Hospital in collaboration with the Department of Neurology.

I first got involved in the analysis of digital EEG as a physics student at the University of Marburg in 1993. As a summer student I ported EEG averaging software written in the 'C' programming language from the Atari/GEM environment to the Vax/VMS platform in collaboration with Prof Frank Rösler. I did my final year project and thesis under the supervision of Prof Frank Rösler and Prof Reinhard Eckhorn on the development of a technique for the decomposition of averaged auditory and somatosensory event-related potentials (ERPs) into a set of spatial components and respective temporal activation signals [Gölz 1995], similar to *independent component analysis* (ICA) [Makeig *et al.* 1996, Ghahremani *et al.* 1996, Makeig *et al.* 1997]. I developed an extensive software package for ERP analysis in the object-oriented 'C++' programming language including a graphical user interface and a script interpreter for psychophysiological research as a research assistant after graduating. Because the ERP waveforms feature a small number of waves with various phase lags, I tried to design a complex wavelet filter with the smallest possible number of oscillations and studied the properties of the resulting CWTs [Gölz *et al.* 1997].

In 1997 I enrolled for a Masters degree in Electrical and Electronic Engineering under the supervision of Dr Phil Bones and Dr Richard Jones and started working on the analysis of clinical EEG and the spike detection problem. In December 1997 I applied for, and was awarded, a University of Canterbury Postdoctoral Scholarship and enrolled for the degree of Doctor of Philosophy.

The development of automated systems for the enhancement and detection of epileptiform spikes in the EEG has a long history as a collaborative project between the Department of Medical Physics and Bioengineering at Christchurch Hospital and the Department of Electrical and Electronic Engineering at the University of Canterbury. One of the earliest spike detection systems was designed by Dr Bruce Davey [Davey *et al.* 1989] and Dr Alison Dingle [Dingle 1992], who applied mimetic feature extraction followed by a rule-based expert system. Dr Christopher James used an artificial neural network (ANN) for the classification of the mimetic features [James 1997] and Dr Donna Ward applied beamformer techniques used in radar systems to the enhancement of deep epileptiform activity [Ward 1998].

There are four main parts in this thesis: part one (Chapters 1 & 2) gives an introduction to the basic anatomy and neurophysiology of the brain, and the processes involved in the generation of electric surface potentials by the activity of the brain known as EEG. An introduction to the various symptoms and seizure types related to epilepsy and to the problem of identifying epileptiform transients that occur between seizures, the spike detection problem, follow.

Part two (Chapters 3 & 4) deals with wavelet-based methods for signal analysis. The two major transforms, DWT and CWT, are presented along with a variety of suitable wavelets. While the DWT is fast and invertible, the CWT is better suited to a detection problem since it is translation-invariant. Transient specific features can be reliably extracted from the CWT only. The wavelet-based approaches are compared to two bilinear time-frequency representations: the Wigner-Ville distribution (WVD) and the Choi-Williams distribution (CWD).

In part three (Chapters 5 & 6) additional methods needed to build a spike detection system are presented. Discriminant functions are used as a tool to classify feature vectors in to two classes. Most discriminant functions have a term called *bias* that balances unequal sample sizes. This is important for the spike detection problem since the number of artifacts is commonly much larger than the number of epileptiform spikes.

Electrostatic models for the generation of the electric surface potential are presented. From a given forward solution, an inverse estimate of the source distribution can be obtained. While the generation of the EEG is confined to the source space, some artifacts may also occur in electrode space. The absence of an absolute zero potential reference makes it impossible to invert the transformation from electrode space to channel space. The effect of an electric current generator in the volume of the source space, the forward solution, is simulated.

In part four (Chapters 7 & 8) a wavelet-based spike detection system is presented, evaluated and compared to two existing systems. The system is based on the CWT and has two stages. In the first stage, transient waveforms are detected with an adaptive multichannel filter based on a single scale of the CWT. In the second stage, multichannel multiscale features of transients are extracted and classified with a linear discriminant function. The results of the new system are compared to those obtained with the two systems developed by Dingle [1992] and James [1997].

During the course of the work presented in this thesis the following workshops, presentations and papers have been prepared:

Goelz, H., Jones, R.D. and Bones, P.J., 'Wavelets and their Application to Analysis of Biomedical Signals' *3rd Annual International Conference of the Australasian College of Physical Scientists and Engineers in Medicine*, Christchurch Hospital, November 1, 1998.

Goelz, H., 'Wavelet Analysis of transient biomedical signals and its application to detection of epileptiform activity in the EEG' *pre-conference Workshop on Time-Frequency and Time-Scale (Wavelets) for biomedical signal and image processing, IEEE Engineering in Medicine and Biology (EMBS) and Bio Medical Engineering Society (BMES) Conference*, Atlanta, October 12, 1999.

Goelz, H., Jones, R.D. and Bones, P.J., 'Continuous Wavelet Transform for the Detection and Classification of Epileptiform Activity in the EEG' *IEEE Engineering in Medicine and Biology (EMBS) and Bio Medical Engineering Society (BMES) Conference*, Atlanta, October 13-16, 1999.

Goelz, H., Jones, R.D. and Bones, P.J., 'Wavelet Analysis of transient biomedical signals and its application to detection of epileptiform activity in the EEG' *Inaugural Meeting of the EEG & Clinical Neuroscience Society (ECNS) at the George Washington School of Medicine*, Washington D.C., October 21-23, 1999.

Goelz, H., Jones, R.D. and Bones, P.J. (2000), 'Wavelet Analysis of Transient Biomedical Signals and Its Application to Detection of Epileptiform Activity in the EEG', *Clinical Electroencephalography* Vol. 41, pp. 181–191.

The following paper has been submitted for publication:

Goelz, H., Jones, R.D. and Bones, P.J., 'A New Family of Fast Wavelet Filters for Transient Detection'.

Chapter 1

BRAIN AND ELECTROENCEPHALOGRAM

1.1 INTRODUCTION

This chapter begins with a brief introduction to the anatomy and neurophysiology of the human brain at various scales. A concise review of the processes involved in the generation of the electric phenomenon related to the brain's activity, the *electroencephalogram* (EEG), is given followed by a presentation of the techniques used in its measurement. Finally, the various normal patterns of the EEG and commonly found artifacts are discussed. Most information in this chapter is drawn from Nolte [1981], Nunez [1981], Duffy *et al.* [1989], Laidlaw *et al.* [1993] and Binnie [1993].

1.2 THE HUMAN BRAIN

1.2.1 Neurons

The basic building elements of the brain are *neurons* and *glial cells.* About 100 billion neurons and ten times as many glial cells make up the human *central nervous system* (CNS). Neurons come in a great variety of sizes and shapes but most have a long protrusion, the *axon*, that carries information from the cell body, the *soma*, to other cells. Special structures at the end of the axons, the *synaptic terminals* connect one cell to another. On the receiving side of the synaptic terminals there are shorter protrusions called *dendrites* that pick up the information and carry it to the receiving cell's soma. Most of the glial cells form a special kind of wrapping, the *myelin sheaths*, around the axons that speed up information conduction and provide electrical insulation.

1.2.2 Principle of Operation

Inside the neuron, information is carried by an electric potential. The neuron's membrane pumps Na^+ out of the soma and K^+ into the soma producing a *resting potential* of about -70 μV. If the neuron 'fires', the membrane transport is rapidly inverted leading to a positive *action potential* that travels down the axon to the synaptic terminals.

Between the neurons, information is carried by chemical processes. A neurotransmitter is released at the synaptic terminals by the firing neuron. The neurotransmitter is picked up by chemical receptors of the receiving neuron's dendrites where it causes a *post-synaptic potential.* The post-synaptic potential may either be *excitatory* or *inhibitory.*

1.2.3 Neuronal Input, Processing and Output

There are three main functional classes of neurons: *sensory neurons*, *interneurons* and *motor neurons.* Sensory neurons collect information from specific nonneural *receptor cells* which are sensitive to visual, auditory, tactile, proprioceptive, olfactory, or other sensory stimuli. Motor neurons directly control muscles and glands. All other neurons are *interneurons* which connect neurons amongst each other and make up 99.99% of the CNS. Interneurons are strongly interconnected with their dendrites having as many as 10^3 to 10^5 synaptic terminals.

1.2.4 Major Structural Elements of the CNS

The CNS is composed of two main components: the brain and the *spinal cord.* The brain can further be subdivided into the *cerebrum*, the *cerebellum* and the *brainstem* (Fig. 1.1). The brainstem is the structure through which nerve fibres relay signals in both directions between the spinal cord and higher brain centres. The cerebellum lies posterior to the brainstem and has a primary role in coordinating movements. The shape of the cerebrum is characterized by a variety of *sulci* (folds), *fissures* (large folds) and *gyri* (rounded protrusions between folds) that are subdivided into four main lobes: frontal lobe, parietal lobe, occipital lobe and temporal lobe as illustrated in Fig. 1.1.

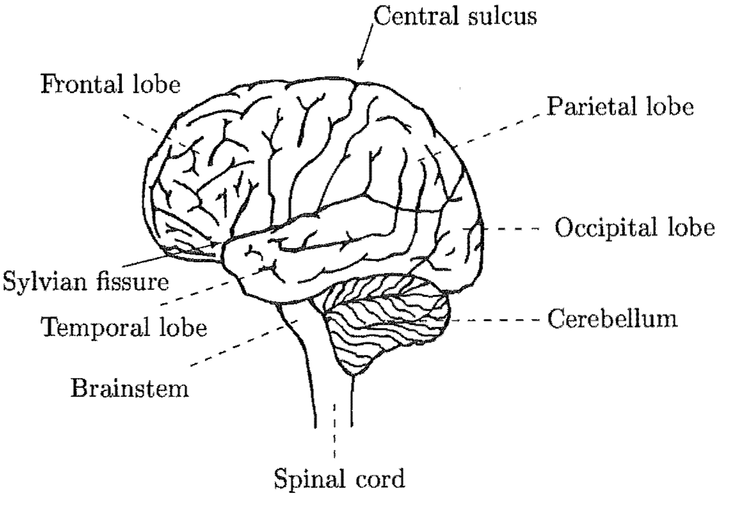

Figure 1.1 Major structures of the cerebrum.

1.2.5 Cerebral Cortex

Of particular interest is the *cerebral cortex*, which is the surface layer of the convoluted structure of the cerebrum. The cerebral cortex is only 2-3 mm thick and can be distinguished as the *gray matter* from the rest of the cerebrum. It contains predominantly cell bodies and dendrites of the neurons whereas the *white matter* is mainly made up of long axons [Nolte 1981].

1.2.6 Pyramidal Cells

The cerebral cortex can be subdivided into six layers (labelled I to VI) each containing characteristic neurons. The largest neurons are the *pyramidal cells* which can grow as large as 100 μm in diameter and occupy layers III and V of the cortex (Fig. 1.2). Pyramidal cells are characterized by a series of *basal dendrites* that extend horizontally from their base, a large vertical *apical dendrite* attached to the top, and a long axon. 95% of the pyramidal cells send out a major axon running into the subcortical white matter that provides a link to a distant cortical area while only 5% make short range intracortical connections [Nunez 1981].

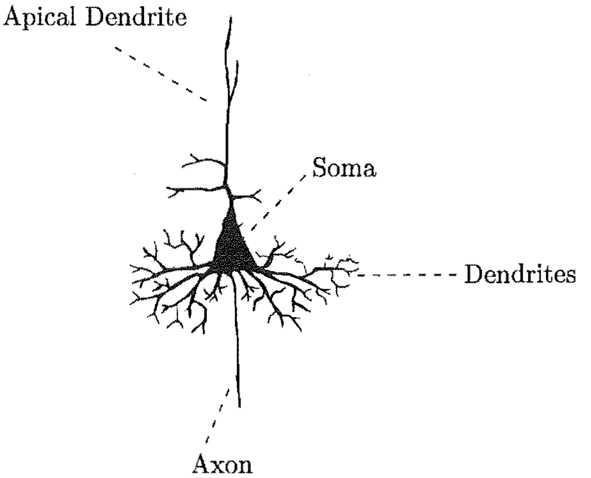

Figure 1.2 Pyramidal cell with soma, axon, basal dendrites and large apical dendrite.

1.2.7 Localization of Function

There is some controversy over localization of function in the cerebral cortex: some researchers maintain that particular areas of cortex are the unique sites of particular functions while others maintain that very large areas of cortex form uniform fields in which functions are not localized and that activities depend on the amount of such cortex that is intact. Although it has been established that damage to particular areas correlates with specific functional deficits, this could be mean that:

- the function is actually localized in a particular area,
- one area performs one crucial step in the function, or
- the area faciliates the activity of one or more other structures [Nolte 1981].

The cortex can be divided into *primary sensory areas*, a *primary motor area* and *association areas*. The location of the primary sensory and motor areas is well known and not disputed (Fig. 1.3). For some areas, such as the visual cortex, many details have

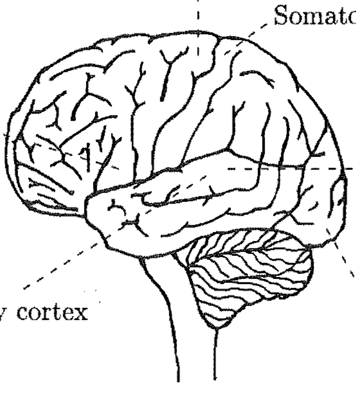

Figure 1.3 Different types of cortex.

been discovered regarding the internal organization and temporal coding of associated signals [Hubel and Wiesel 1977, Eckhorn *et al.* 1988, Herculano Houzel *et al.* 1999]. Only two areas representing higher cognitive functions are well established: *Broca's area* and *Wernicke's area*, which relate to the production and comprehension of spoken language. Both areas have been studied in relation to *aphasia* (loss of speech) induced by lesions. Broca's area is involved in speech production while Wernicke's area can be linked to the comprehension and production of language (Fig. 1.3).

A comprehensive collection of studies covering various modalities such as *x-ray computed tomography* (CT), *positron emission tomography* (PET), *functional magnetic resonance imaging* (fMRI), *electroencephalography* (EEG), *magnetoencephalography* (MEG), in investigations of the localization of brain function has recently been published by Toga and Mazziotta [2000].

1.3 THE ELECTROENCEPHALOGRAM

1.3.1 Origin

Neurons undergo two main types of electrical change: slow postsynaptic potentials lasting for 5-50 ms and much briefer action potentials lasting only about 1 ms. Action potentials do not generate fields that are recordable far from their origin. Virtually only postsynaptic potentials contribute to the field as measured on the scalp as the

EEG. The peak amplitude of the EEG is in the range of $5\,\mu$V to $100\,\mu$V, but it is mostly below $50\,\mu$V [Duffy *et al.* 1989]. The activity of the pyramidal cells with their large apical dendrites is believed to be the major contributor to the generation of the EEG. Only the gross spatial mean of electrical activity of the cells can be recorded on the scalp. Binnie [1993] suggested the following analogy: someone listening outside a rugby stadium may hear but cannot understand the confused shouts of the crowd for most of the time. However, the general applause when the home team scores a try and the universal gasp when one is missed are easily recognized. That is, EEG recordings primarily reflect only that neuronal activity which is synchronous over a large number of neurons.

A complete model of the physiological basis of the EEG, linking the microscopic level of neurons and both their local and global interconnection properties with the resulting potentials at a macroscopic level was proposed by Nunez [1981]. The basic mechanism generating surface potentials is illustrated in Fig. 1.4 [Nunez 1995].

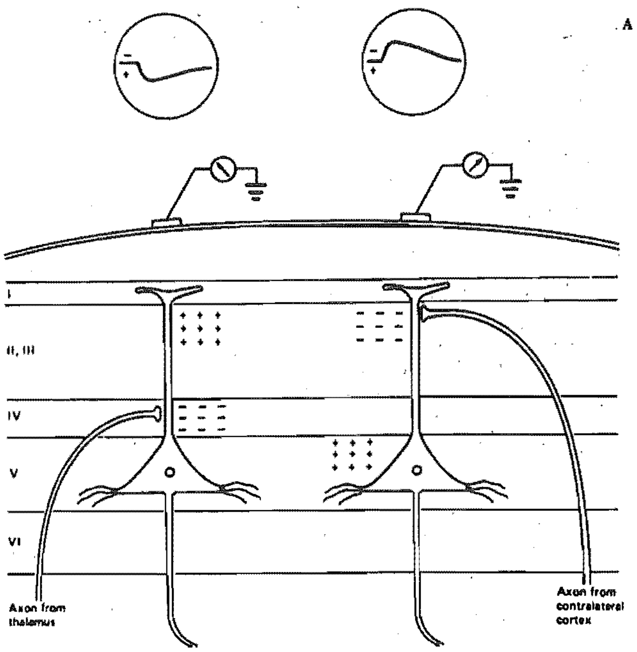

Figure 1.4 The six cortical layers (I-IV) with two pyramidal cells in layer V. Postsynaptic potentials generate surface positivity if the synaptic terminals are located in layer IV. Surface negativity results from synaptic terminals in layer II [Nunez 1995].

1.3.1.1 Rhythmic Activity

The most common rhythmic electrical brain activity is found near 10 Hz. It is known as the *alpha rhythm* and was reported by Berger [1929].

Andersen and Andersson [1968] found that the *thalamus*, a small, olive-shaped cerebral structure above the brainstem, has a putative role as a pacemaker for the alpha rhythm. Besides its presence in the EEG, the rhythm can be recorded with microelectrodes in the thalamus. After removal of the thalamocortical connections the alpha rhythm was found in the thalamus only [Andersen and Andersson 1968].

Lippold [1973] suggested that alpha waves were generated by the tremor in the extra-ocular muscles rather than the brain. However, his suggestion was rejected by the majority of scientists.

Nunez [1981] reported that the phase relationships between thalamic and cortical alpha waves were far less consistent than those between distant cortical areas. He suggested a model for the intracortical generation of the alpha rhythm which is based on the wave equations used in physics to describe properties of standing and moving waves. The interaction between thalamic and cortical alpha waves may be more complex than suggested by Andersen and Andersson [1968].

1.3.2 Recording System

Electric potentials on the scalp are commonly picked up by 20 or so electrodes. In routine clinical recordings, electrodes are placed at well defined locations according to the international 10-20 system proposed by Jasper [1958]. Four fixed landmarks serve as a references: nasion, inion, and the two preauricular points (Fig. 1.5). The 10-20

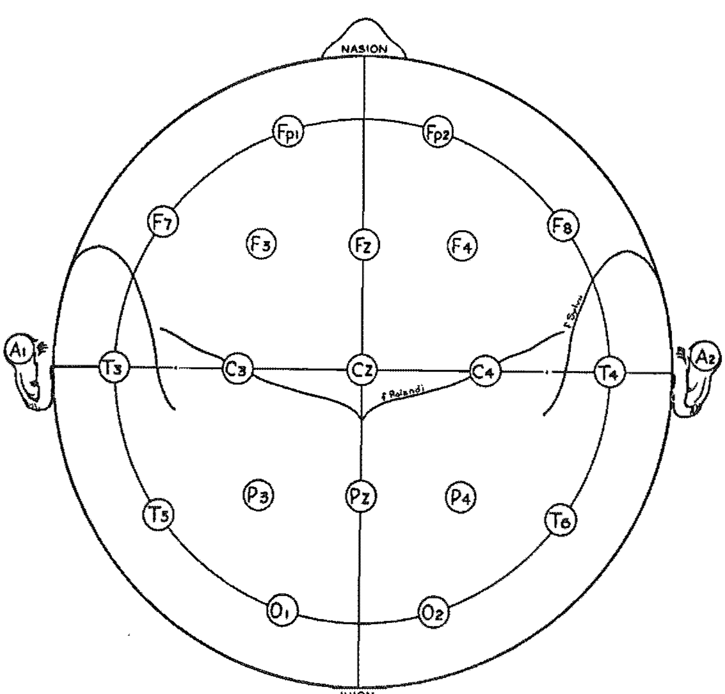

Figure 1.5 System for electrode placement proposed by Jasper [1958].

system divides the perimeter of the head into sectors of 10 and 20% (e.g. perimeter through nasion-Fp1-F3-C3-P3-O1-inion). Potentials are measured either with respect to a single reference electrode (referential recording) or with respect to an adjacent electrode (bipolar recording).

The manner in which electrodes are mapped to channels is called the *montage* and may be changed several times within a recording session. Recordings are preferably made within an electrically shielded room. The EEG signals in each channel are amplified and band-pass filtered (commonly 1-70 Hz). Traditionally, recordings of 8, 16 or 21 channels were instantly written by a bank of galvanometers on a lengthy paper chart but this has largely been replaced by paperless recording systems in which EEGs are digitized, stored, reviewed and archived by computer-based systems. Digital EEG has several fundamental advantages including the ease of changing the filter and montage of an actual recording. Montages used in the current work are listed in Fig. 6.1.

1.3.3 Features of Brain Waves

A variety of rhythms and transient waveforms make up the normal *background* or *ongoing* EEG. Features of the background EEG are mainly influenced by the age of the subject and the state of sleep or wakefulness. Features of normal and abnormal EEG patterns are described in more detail in Duffy *et al.* [1989] and Fisch [1991].

1.3.3.1 Alpha Activity

The most prominent normal feature from the adult human cortex are *alpha waves* or *alpha spindles* which are characterized by a regular series of waves at about 10 Hz. Alpha

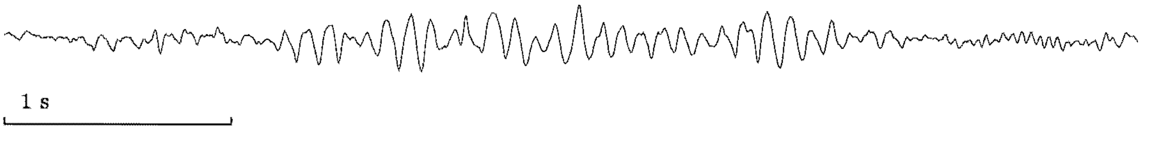

Figure 1.6 Alpha spindles.

waves between 8.5 and 12 Hz are considered normal for adults. Alpha activity may originate from all regions but is more characteristic of the occipital areas [Ogilvie 1949]. It is most obvious while subjects are resting with eyes closed and diminishes dramatically when eyes are opened. It may show a waxing and waning pattern over time. An example of alpha waves is given in Fig. 1.6 and in Fig. A.11 on p. 166 in Appendix A.

Kellaway [1979] found that the mean frequency of alpha waves rises from 8.0 Hz in newborns to 10.0 Hz at the age of 22 years. In a similar study, Sem-Jacoben [1953] report a drop of the mean alpha frequency from 10.3 Hz in 17 to 29-year-olds to 9.2 Hz in those over 60.

1.3.3.2 Beta Activity

The *beta rhythm* is of lower amplitude than alpha activity and lies within the frequency band 18 to 32 Hz. It is more dominant in the anterior regions of the head and may be present simultaneously with alpha activity. An illustration of beta activity is shown in Fig. 1.7 and in Fig. A.3 on p. 158 in Appendix A.

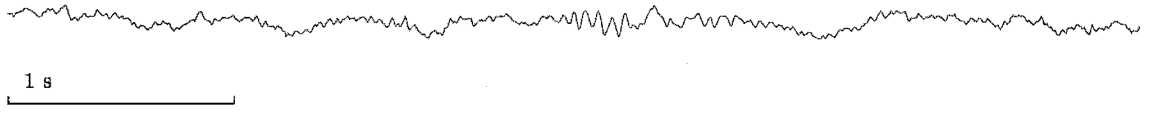

Figure 1.7 Beta activity.

1.3.3.3 Theta Activity

Theta waves lie within the frequency band of 4-7 Hz and occur during light sleep in normal adults. Theta waves found in young children during wakefulness may represent alpha activity of low frequency which can only be distinguished by recording periods with eyes open and eyes closed [Kellaway 1979].

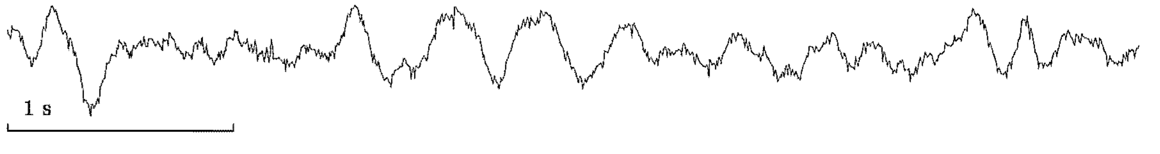

Figure 1.8 Theta waves.

1.3.3.4 Delta Activity

Delta activity comprises slow waves of 1-4 Hz and are only found in sleep stage IV of normal subjects. Excess delta activity is also a strong indicator of structural lesions in the brain such as tumors or haemorrhage.

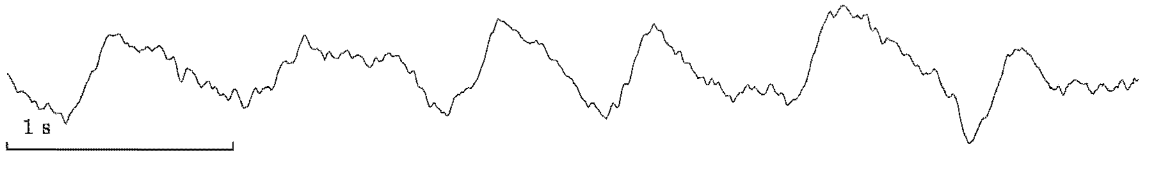

Figure 1.9 Delta waves.

1.3.3.5 Gamma Activity

Cortical oscillations between 35 and 55 Hz are known as *gamma activity.* These waves are usually of extremely small amplitude and are often not seen at all [Ogilvie 1949]. Coherent stimulus-evoked resonances between 35 and 75 Hz have been reported from the visual cortex during the presentation of a simple visual stimulus [Eckhorn *et al.* 1988]. Tallon-Baudry *et al.* [1996] present less coherent visually evoked gamma-band responses in the EEG.

1.3.4 Artifacts

There are frequently other rhythms and transient waveforms in the EEG which are of non-cortical origin. In some cases, considerable experience is required to tell artifacts from cortical activity.

1.3.4.1 Muscle Activity

Artifactual muscle activity is often seen in EEG with jaw movements or contraction of facial muscles. It may be continuously present if the subject is unable to relax. The electrical potentials resulting from muscle contractions appear as fast spiky waves at a rate of about 30 to 60 Hz [Ogilvie 1949].

1.3.4.2 Sweating

When the subject perspires, the baseline of the tracing may rise and fall, at a rate of 1 Hz or less. The movements are more prominent in the frontal leads and are thought to be caused by the activity of the sweat glands [Ogilvie 1949].

1.3.4.3 Drift Artifacts

Perspiration and other electrochemical processes at the skin-electrode interface may cause a continuous quasi-linear channel-specific drift. The drift is only observed in DC-recordings which are commonly made to explore slow event-related activation patterns. It can be removed by linear regression [Hennighausen *et al.* 1993].

1.3.4.4 Electrode Artifact

Electrodes may suddenly lose contact or become high resistive, producing abrupt jagged patterns also called 'electrode pop'.

1.3.4.5 ECG Artifact

A series of isolated waves occurring at a rate of one to two per second may be caused by the QRS component of the *electrocardiogram* (ECG). The artifact is mainly found in referential channels. ECG artifacts are illustrated in Fig. A.17 on p. 172 in Appendix A.

1.3.4.6 Eye Blinks and Eye Movements

A constant potential exists in the retina such that eye movements cause strong potential changes on the scalp around the eye called the *electrooculogram* (EOG). The EOG can be recorded with bipolar channels near the eye. Eye blinks cause large scale positive transient deflections in the frontal channels of the EEG. A series of eye-blinks in the EEG can be seen on p. 174 in Appendix A. Artifacts caused by eye-blinks are usually symmetric on the two sides of the head and show decreasing amplitude from the frontal to the rear channels. However, the spatial distribution may change with the direction of gaze of the subject. Horizontal and vertical eye movements may also cause small low frequency transients in frontal channels. Fast eye blinks are especially problematic in the detection of epileptiform transients, since they exhibit some similarity.

1.3.4.7 Electric Appliance Artifact

The ring of a telephone in the EEG laboratory may sometimes cause a series of 18-25 Hz waves in the EEG [Ogilvie 1949]. It may be a difficult and time-consuming task to identify the source of any enviromental electric disturbance.

1.3.4.8 Mains Interference

A steady 50 or 60 Hz artifact may be induced from the mains alternating current. Many EEG systems are equipped with a notch filter that strongly attenuates the respective frequency. Mains interference may be caused by loops in the ground wiring.

1.4 SUMMARY

The basic anatomic structures and physiological processes of the human brain that are involved in the generation of the EEG were presented. Common features of the normal background EEG and various artifacts were described and illustrated.

Chapter 2

EPILEPSY AND THE SPIKE DETECTION PROBLEM

2.1 INTRODUCTION

Important aspects of epilepsy and the EEG as a diagnostic tool are presented. Changes in the EEG during and in between seizures are described.

Various approaches to the automatic detection of interictal epileptiform activity in the EEG are presented. Approaches are based on mimetic features, statistical (autoregressive) models or transforms such as the wavelet transform.

2.2 EPILEPSY

Epilepsy is a reasonably common neurological disorder. It is estimated that about 1% of the world's population has epilepsy and, despite medication, about 10% of patients have seizures more than once a month. As epilepsy is a symptom complex and not a disease as such, a presentation of epilepsy requires reference to three levels: the aetiology (cause), the seizure type, and the epileptic syndrome [Porter 1993].

The various classifications of epileptic seizures and epileptic syndromes are still evolving, but two standards have been approved by the International League Against Epilepsy: the 1981 Classification of Epileptic Seizures and the 1989 Classification of Epileptic Syndromes [Comm. ILAE 1981 and 1989]. Epilepsy can be caused by virtually any major category of serious disease or disorder in humans. It can result from congenital malformations, infections, tumours, vascular diseases, degenerative diseases, or injury. If the cause can be identified, the epileptic syndrome is labelled 'symptomatic', otherwise - or until the cause has been identified - it is labelled 'cryptogenic' [Porter 1993].

2.2.1 Epileptic Seizures

Epileptic seizures are the most important indicator for an epileptic syndrome. Therefore it is of major importance to make a clinical observation of a seizure in order to diagnose an epileptic syndrome. Long-term observation of a patient with simultaneous video and

EEG recording may be needed to provide documentation of seizures that occur rarely. If no clinical observation can be obtained the description given by eye witnesses may be needed, particularly for those types of seizures that involve loss of consciousness.

Epileptic seizures are classified as *partial seizures* or *generalized seizures.* Partial seizures are those in which the first clinical and EEG changes indicate initial activation of a system of neurons in one cerebral hemisphere. If both hemispheres are involved in the initial neuronal activation, the seizure is labelled 'generalized'. *Simple partial seizures* do not impair consciousness, *complex partial seizures* do impair consciousness and *secondarily generalized partial seizures* evolve to generalized seizures. Types of partial and generalized seizures and associated clinical signs are listed in Table 2.1.

Seizures may last from seconds to several minutes. Seizures lasting for more than 30 minutes or occurring at such frequency that the patient is unable to recover to a normal level of functioning between seizures are called *status epilepticus* [Treiman 1993]. For the diagnosis of an epileptic syndrome, additional factors such as case history, family history, age, age at onset, gender, neurological examination, ictal/interictal EEG (see below), psychological evaluation, aetiology of epilepsy, angiogram, MRI scan, CT scan, or the examination of the cerebrospinal fluid may play an important role beside the type of seizure observed [Porter 1993].

2.2.2 Changes in the EEG

2.2.3 Ictal Changes

The EEG recorded during a seizure is called *ictal EEG*, while the EEG between seizures is referred to as *interictal EEG.* Most seizures produce clear signs in the ictal EEG that are valuable indicators for the diagnosis. For example, the characteristic pattern found in ictal EEG of patients having absence seizures are 2-4 Hz spike-and-wave complexes as shown in Fig. A.1 on p. 156 in Appendix A.

2.2.4 Interictal Changes

In the interictal EEG, characteristic patterns for epilepsy may be less obvious. Interictal epileptiform patterns may include *spikes*, *spike-and-wave complexes*, and *sharp waves.* Fast abnormal transients such as the epileptiform patterns are called *paroxysmal activity.*

Chatrian *et al.* [1977] defined *epileptiform transients* (ETs) on behalf of the International Federation of Societies for Electroencephalography and Clinical Neurophysiology (IFSECN) as 'transient, clearly distinguished from background activity, with pointed peak at conventional paper speeds and a duration from 20 to under 70 msec, that is, 1/50th to 1/13th of a second approximately'.

Binnie [1993] defined epileptiform transients as follows: 'spiky transients which are of less then 80 ms duration are called spikes, those lasting from 80-120 ms are

Table 2.1 Major seizure types and associated changes in the EEG. Up to six sub-categories with individual clinical symptoms may exist for each clinical seizure type [Commission on Classification and Terminology of the International League Against Epilepsy 1981].

Clinical seizure type	EEG seizure type	Interictal EEG
	Partial Seizures	
Simple partial seizures (consciousness not impaired)	Local contralateral discharge starting over the corresponding area of cortical representation (not always recorded on the scalp)	Local contralateral discharge
Complex partial seizures (with impairment of consciousness; may sometimes begin with simple symptomatology)	Unilateral or, frequently bilateral discharge, diffuse or focal in temporal or frontotemporal regions	Unilateral or bilateral generally asynchronous focus; usually in the temporal or frontal regions
Partial evolving to secondarily generalized seizures (This may be generalized tonic-clonic, tonic, or clonic)	Above discharges become secondarily and rapidly generalized	
	Generalized Seizures	
Absence or 'petit mal' seizures	Usually regular and symmetrical 3 Hz but may be 2-4 Hz spike and slow-wave complexes. Abnormalities are bilateral	Background activity usually normal although paroxysmal activity (such as spikes or spike and slow-wave complexes) may occur. This activity is usually regular and symmetrical
Atypical absence	EEG more heterogeneous; may include irregular spike and slow-wave complexes, fast activity or other paroxsymal activity. Abnormalities are bilateral but often irregular and asymmetrical	Background usually abnormal; paroxysmal activity (such as spikes or spike and slow-wave complexes) frequently irregular and asymmetrical
Myoclonic seizures Myoclonic jerks (single or multiple)	Polyspike and wave, or sometimes spike and wave or sharp and slow waves	Same as ictal
Clonic seizures	Fast activity (10 c/s or more) and slow waves; occasional spike and wave patterns	Spike and wave or polyspike-and-wave discharges
Tonic seizures	Low voltage, fast activity or a fast rhythm of 0-10 c/s or more decreasing in frequency and increasing in amplitude	More or less rhythmic discharges of sharp and slow waves, sometimes unsymmetrical. Background is often abnormal for age
Tonic-clonic seizures	Rhythm at 0-10 or more c/s decreasing in frequency and increasing in amplitude during tonic phase, interrupted by slow waves during clonic phase	Polyspike and waves or spike and wave, or, sometimes sharp and slow wave discharges
Atonic seizures	Polyspikes and wave or flattening or low-voltage fast activity	Polyspike and slow wave

sharp waves. Waves of longer duration than 120 ms and of sharp appearance may also represent epileptiform activity. Discrete spikes and sharp waves are usually followed by a slower wave and, if this is prominent, the two together are described as spike and wave complex. Focal spikes are often polyphasic, but the most prominent component is usually surface negative' (see below).

2.2.5 Focal and Non-focal Epileptiform Discharges

An epileptiform event causes a number of coincident transients such as spikes, sharp waves, or spike-and-wave complexes in several EEG channels. Events on single channels are deemed to be ETs while the set of all EEG multichannel transients related to a single event is called an *epileptiform discharge* (ED).

EDs emerging from a well localized region of cerebral irritation such as that caused by tumours or injuries are called *focal EDs.* If several foci are found in a single recording, the recording is said to contain *multifocal* epileptiform activity.

A focal ED shows several peaks in adjacent bipolar channels. The specific property of a *surface negative* focal ED is that the polarity of the peaks changes such that peaks on one pair of adjacent channels point at each other. This phenomenon is known as *phase reversal.* In some cases the pair of channels that show the phase reversal may be separated by a channel not showing any unusual activity, a *null-channel.* A surface negative focal ED without null-channel is shown in Fig. A.20 on p. 175. Several focal EDs from multiple foci are illustrated in Fig. A.6 on p. 161.

Other EDs that do not emerge from a well localized area of the cerebrum are called *generalized EDs.*

The EEG samples in Appendix A show a variety of interictal epileptiform patterns as described in the captions.

2.2.6 Diagnostic limitations of the interictal EEG

2.2.6.1 EEG Spikes in Healthy Subjects

Binnie [1993] has listed several examples of spike-and-wave complexes found in the EEG of people who do not suffer from epilepsy.

One of the examples refers to rhythmic spiky discharges at 6 and 14 Hz found in drowsiness and light sleep of young adults. Spikes were surface-positive as opposed to epileptic spikes. About 20% of cases studied showed this feature.

Mid-temporal rhythmic discharges consist of rhythmic sharp waves at about 6 Hz over the temporal regions. This activity appears to be unaffected by vigilance, attention, hyperventilation, and is unresponsive to epileptic drugs. During the discharge no ictal clinical manifestations are seen and no cognitive disturbance is demonstrable.

A similar phenomenon, *sub-clinical rhythmic discharges of adults* (SREDA), occurs further posteriorly and is virtually confined to adults of more than 50 years of age. It appears to have no clinical significance either. Chadwick [1993] expressed the function and limits of the diagnostic value of the EEG in the following way:

> *"The EEG provides valuable information that may:*
> *(a) add weight to the clinical diagnosis;*
> *(b) aid the classification of epilepsy, and*
> *(c) show changes that may increase the suspicion of a structural lesion.*
> *Routine interictal EEG recording is one of the most abused investigations in clinical medicine and is unquestionably responsible for great human suffering. The diagnostic value of an interictal EEG is widely misunderstood. EEGs are often requested either to exclude or to prove a diagnosis of epilepsy - something that can seldom, if ever, be done. Erroneous interpretation of the EEG is probably the commonest reason for non-epileptic events being diagnosed as seizures."*

2.2.6.2 Epileptic Patients without EEG Spikes

Conversely, in a study by Binnie *et al.* [1986], 50% of 100 patients with known epilepsy showed no signed of interictal epileptiform activity in the initial clinical EEG recording. When followed by sleep EEGs, 80% showed interictal activity. In 8 of 100 epileptic patients no signs of interictal epileptiform activity could be found in repeated wake and sleep recordings [Binnie *et al.* 1986].

2.3 AUTOMATIC DETECTION OF INTERICTAL SPIKES IN THE EEG

MacGillivray [1977] notes that neurophysiologists are fairly efficient at detecting spikes in the EEG within their own terms of reference and that repeatability for a single reader is high. However, inter-rater agreement can be surprisingly low. The precise definition of the sequence of operations by which marking is done in all circumstances, as is required for computer emulation, is a very difficult task and can only be achieved approximately.

Glover *et al.* [1986] summarize the advantages automated methods of EEG scoring may have over visual methods:

> *(a) They can ease the work-load of the EEGer by providing off-line, faster than real-time analysis of lengthy records*
> *(b) They can provide reliability and repeatability in the analysis of data*

(c) They can offer a tool for detailed quantification of the epileptiform activity, which could be used to study the effect of drug treatment
(d) They could eventually lead to a comprehensive definition of an epileptiform, and thus contribute to the standardization of epileptiform detection.

Two primary areas of application were identified by Dingle *et al.* [1993]: firstly, long-term EEG monitoring, where automated detection of epileptiform transients acts as a data reduction process, and, secondly, in routine clinical recordings where the objective is to minimize the visual inspection required.

2.3.1 The Spike Detection Problem

Various approaches have been taken to automatically identify epileptiform spikes. EDs, background features and artifacts come in a large variety. Furthermore, background features can change rapidly over time. All patterns found in the EEG have individual durations. Thus, an appropriate amount of temporal context is needed to identify them. The *spatial context* may provide an important clue as EDs lead to a pool of surface negativity. Patterns may be of short duration but recurrent such that their detection can only be verified by comparing distant epochs in a recording or, in other words, *wide-sense temporal context* [Dingle *et al.* 1993, Black *et al.* 2000].

A number of approaches to the spike detection problem are presented in the following subsections. The performance of some of these systems is presented in Chapter 8.

2.3.2 Second Derivative Techniques

Carrie [1972b] calculated the second derivative of EEG waves on an analogue computer. When the magnitude of the second derivative exceeded a threshold, a transient was considered to be detected. No further classification of the transient was made. The threshold was dynamically updated with respect to a predefined number of previous waves. Waves were delimited by zero crossings of the EEG signal. Only a single EEG channel was processed and no performance evaluation was made [Carrie 1972b, Carrie 1972a].

2.3.3 Zero Crossing Techniques

Ehrenberg and Penry [1976] counted the zero crossings in two mean EEG signals derived from the average of four channels. Running sums of amplitudes were decreased by 12.5% per zero crossing unless the wave duration was within 50-71 ms (14 to 20 cycles per seconds, respectively). If the running sums exceeded a threshold for more than 750 ms a generalized spike-and-wave complex was considered to have been detected. Only one kind of ET was detected by the system. Multichannel context was not used and the performance of the system was not evaluated.

2.3.4 Mimetic Approaches

Mimetic approaches try to mimic the way EEG readers evaluate individual waves. Firstly, the single channel EEG is broken down into a series of waves and half-waves. Subsequently, several characteristic attributes are assigned to each wave.

Kooi [1966] evaluated the amplitude, duration, velocity and the angle subtended by the rising and falling slope of various ETs. He suggested using these features to define objective criteria of ETs in paper recordings before the first spike detection systems were built. The angle definition he used is illustrated in Fig. 2.1 (left) with an actual ET. The respective ED can be found in channel 16 in Fig. A.6 on p. 161. Gotman and

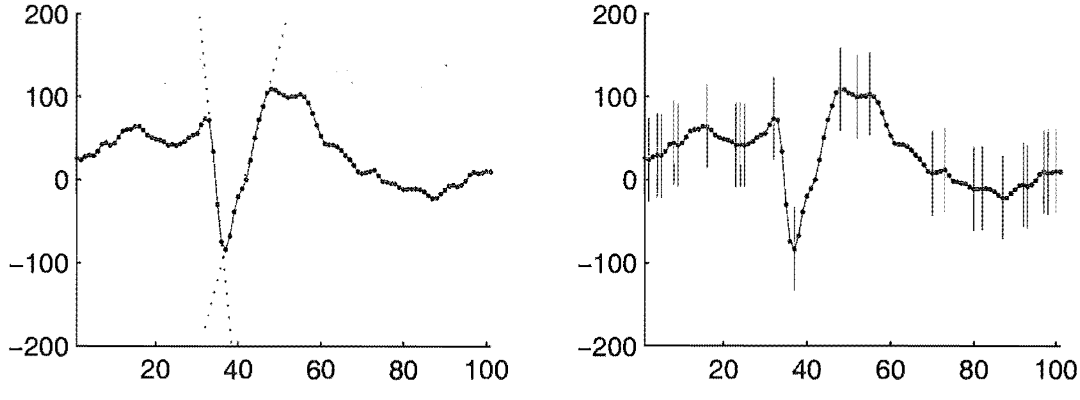

Figure 2.1 Sample ET (100 samples, 0.5 s from E8219). Left: linear regression of rising and falling slopes of the spike. The angle between the two lines was proposed as a feature by Kooi [1966]. Right: periods of rising and falling potential between local extrema. Periods were joined to longer sequences in subsequent processing stages [Gotman *et al.* 1976]

Gloor [1976] split EEG signals into rising and falling segments as illustrated in Fig. 2.1 (right). Segments were subsequently joined to larger elements they called 'sequences' and 'waves'. Several basic attributes were assigned to each wave:

- **relative amplitude:** amplitude of waves normalized by the average of the preceding five seconds,
- **duration of half-waves:** duration between minimum and maximum,
- **pseudo-duration of half-waves:** duration with non-linear correction (see below),
- **relative sharpness:** second derivative of the apex normalized by preceding five seconds, and
- **overall duration of the wave.**

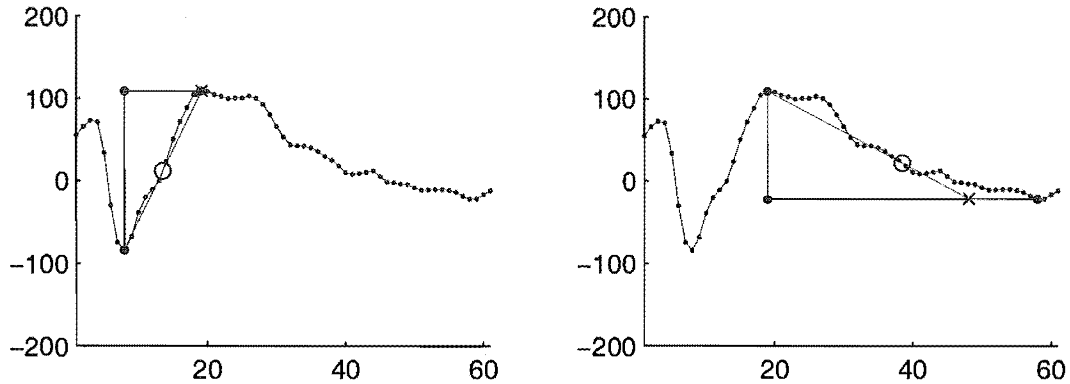

Figure 2.2 Close up of sample ET in previous figure showing the three major half-waves of the spike (60 samples only). For the second and third half-waves amplitude and duration are indicated by vertical and horizontal lines, respectively. Circles mark the potential at half the duration. Crosses (×) mark pseudo-duration as suggested by Gotman *et al.* [1976]. Two cases are illustrated: (left) the pseudo-duration equals the duration, and (right) the pseudo-duration is smaller than the duration.

The amplitude, duration and pseudo-duration of half-waves, as defined by Gotman and Gloor [1976], are illustrated in Fig. 2.2. Subsequent processing rejected muscle spikes, eye-blinks and alpha activity. Inter-channel relationships were analyzed with respect to phase reversal patterns in bipolar chains of channels to confirm surface negativity. Similar mimetic approaches were used by Gotman [1985], Glover *et al.* [1986], Glover *et al.* [1989], and Davey *et al.* [1989]. Dingle *et al.* [1993] measured the amplitude of a wave with respect to a 75 ms floating mean centered around the peak. They used the sum of the magnitudes of the half-waves' slopes as a measure of sharpness [Dingle *et al.* 1993]. Performance of the system was tested in a clinical study with 521 EEGs including 50 epileptiform EEGs [Jones *et al.* 1996]. For the classification of entire EEGs a sensitivity of 95% and a selectivity of 72% were reported. There were only 0.29 false detections per hour.

2.3.5 Parametric Approaches

Da Silva *et al.* [1975] applied an autoregressive model to the EEG. Non-stationarities such as epileptiform spikes were identified by a large prediction error of the model. They assumed a chi-square distribution for the prediction error and derived detection thresholds for confidence levels of 0.5% and 0.1%. Birkemeier *et al.* [1978] used the second derivative of the EEG and the second derivative of the prediction error of an autoregressive model as transient indicators.

2.3.6 Syntactic Approach

Walters *et al.* [1989] assigned one out of five symbols to each half wave according to slope and duration. Thus, they derived a symbolic representation of ETs from the corresponding sequence of symbols they called 'spike string'. A finite-state grammar was automatically inferred from a set of training strings. The finite-state grammar represented a class of strings larger than the training set with properties of ETs.

2.3.7 Approaches based on Artificial Neural Networks

Eberhart *et al.* [1989] trained an *artificial neural network* (ANN) with 240-ms windows (60 samples) of ETs and non-ETs from a single-channel EEG recording. They examined several possible neural network architectures.

Webber *et al.* [1994] developed a spike detection system based on a 3-layer feed-forward ANN. The system was trained with the error-backpropagation method. The sensitivity at equal specificity of the system was increased from 46% to 73% by using mimetic features of ETs rather than 'raw' EEG data (refer to Chapter 5 for definitions of sensitivity and specificity). A 3-layer feed-forward ANN was also used by Gabor and Seyal [1992].

Kalayci and Özdamar [1995] trained an ANN with features derived from wavelet analysis. Their approach is detailed in section 2.3.9.

James [1997] used mimetic feature extraction and a *self-organizing feature map* (SOFM) for single channel analysis. The spatial context was subsequently evaluated by a fuzzy logic stage [James 1997, James *et al.* 1996a, James *et al.* 1996b, James *et al.* 1999]. After training with 35 epileptiform EEGs, a sensitivity of 55% and a selectivity of 82% were achieved in a blind test with 7 epileptiform and one normal recordings. James *et al.* [1997] also applied *multireference adaptive noise cancelling* (MRANC) to the raw EEG which they proposed as a spatial pre-processing stage to improve the sensitivity of any spike detection system.

Tarassenko *et al.* [1998] compared the performance of two sets of features, mimetic features and coefficients derived from an autoregressive model (prediction error and partial correlations), in the training of an ANN. Superior performance was reported for the autoregressive model coefficients. However, a combination of all features further improved the classification to a mean of 93% sensitivity and 91% specificity in a blind cross-validation of recordings from four patients.

Ko and Chung [2000] trained an ANN with 6 hidden layers with 'raw' ETs and non-ETs. They reported a sensitivity of 87% at equal specificity.

2.3.8 Rule-based Systems

Glover *et al.* [1989] and Dingle *et al.* [1993] combined the mimetic feature extraction with a rule-based expert system for classification. A rule-based expert system can verify or falsify a fact based on knowledge present in a number of rules. Specific *declarative programming languages* such as *Lisp* (short for 'list processor') or *Prolog* ('programming in logic') support the representation of knowledge [Charniak and McDermott 1985]. Other than common *procedural programming languages*, Prolog can deduce new facts from given knowledge and automatically find strategies to verify or falsify a fact in question. Ramabhadran *et al.* [1999] improved the original system of Glover *et al.* [1989] and achieved a sensitivity of 95.7 % and a selectivity of 88.9%[1] with rule-based classification of mimetic features.

2.3.9 Wavelet-Based Approaches

In order to locate ETs in an EEG signal it is reasonable to use a filter that approximates a single spike like a template. If the filter is chosen such that it also features zero mean and good localization in the frequency domain it can be considered to be a wavelet filter. Wavelet-based approaches make use of either the *continuous wavelet transform* (CWT) or the *discrete wavelet transform* (DWT).

Approaches presented in this section are based on wavelet methods introduced in the following chapter.

The Mexican hat wavelet (Fig. 2.3(a)) was used by Schiff *et al.* [1994b] to analyze

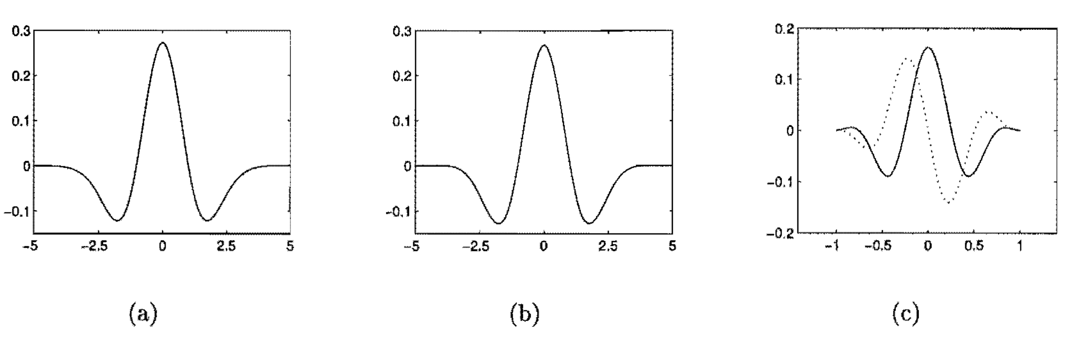

Figure 2.3 (a) Mexican hat wavelet used by Schiff *et al* [1994a], (b) cubic spline approximation of Mexican hat wavelet used by Schiff *et al* [1994b] and (c) wavelet proposed by Senhadji *et al* [1995]. Dotted line shows imaginary part.

data from patients with implanted subdural electrodes. Several spikes and the onset of a seizure could be seen in a series of transforms they calculated using dyadic scales and translations. The same events could not be identified in Fourier spectrograms using various window lengths. A reasonable approximation to the Mexican hat function was acquired by cubic spline interpolation of three points (maximum and both minima of the original wavelet $\psi(t)$, Fig. 2.3(b)). This interpolation allowed the application of a faster algorithm while achieving the same results. Processing time went down from 3 days on a Sun workstation to 3-25 s on a Macintosh. They used a 'random surrogate data set' which had the same power spectrum as the original signal in order to define

[1]Ramabhadran *et al.* report a normalized false detection rate (FD)/(DT + FD) of 11.1% which equals 88.9% selectivity. Refer to section 5.3.1 for details on performance measures.

the threshold for non-random events in each scale at 1.96 standard deviations (ca. 95% confidence level, assuming a Gaussian distribution) above the average.

Schiff *et al.* [1994a] also used a two-dimensional variation of the Mexican hat wavelet as a spatial filter to localize the area of seizure onset. Since subdural electrodes were arranged in a regular 5 by 5 grid, the 2D Mexican hat wavelet could be employed as a spatial centre-surround filter on that grid, a procedure similar to estimation the radial *current source density* (CSD) with the surface Laplacian [Perrin *et al.* 1987] (see Chapter 6 for more details of the spatial analysis). Both methods give a more localized measure of brain activity than the original potential recordings. In contrast to the CSD, the wavelet method allows the observation of different scales. Three scales of the 2D Mexican hat wavelet were suggested and the smallest scale was finally applied. The electrodes that corresponded to the onset of a clinical seizure were readily identified with this method.

Mehta *et al.* [1994] examined the variances of wavelet coefficients of three data sets corresponding to normal activity, seizure onset and ongoing seizure. During normal activity a scale-invariant property of the variances was observed indicating an underlying $1/f$ or $1/f^{\alpha}$ process. This feature disappeared during seizure activity. They proposed monitoring variations in the variance structure of the wavelet coefficients to detect the onset of seizure activity.

Kalayci and Özdamar [1995] used the DWT of clinical EEG recordings containing spike and slow wave complexes to reduce the input size of an ANN within a multistage system for spike detection. Transient waveforms were classified by to two EEGers. The Daubechies-4 and Daubechies-20 wavelets were used in the DWT of a 512 sample (5.12 s) window centered around the transients [Daubechies 1988]. Eight wavelet coefficients of single scales and combinations of scales were used for training (1200 data sets) and testing (2414 data sets) the ANN. They interpreted their results as a 'drastic decrease of the input size of the ANN, without much compromise in its performance'. The best (not the average) resulting sensitivities and selectivities are listed and compared to the performance achieved with 20 (or 30) raw samples used for training the ANN. The results appear, however, to depend rather on the input size itself than on the application of the wavelet preprocessing proposed [Kalayci *et al.* 1994, Kalayci and Özdamar 1995].

Senhadji *et al.* [1995] applied the CWT to detect EDs in EEG recordings and proposed the wavelet $\psi(t) = (1 + \cos(2\pi f_0 t)) \cdot e^{i2\pi k f_0 t}$ with $|t| \leq \frac{1}{2f_0}$ and integer $k \neq \{-1, 0, 1\}$. Fig. 2.3(c) shows this wavelet for $k = 2$, the smallest positive value and scaling factor $f_0 = 1$. They compare a weighted sum of the squared coefficients obtained by the CWT for a predefined set of scales to a threshold λ_1 to detect transients in all channels:

$$\sum_{i=1}^{n} \alpha_i |D_{a,n}|^2 > \lambda_1$$

(a) (b)

Figure 2.4 (a) Morlet's wavelet ($\omega_0 = 5$) used by Clarencon *et al.* [1996] and (b) Polynomial wavelet used by D'Attelis *et al.* [1997].

with weight factors α_i and wavelet coefficients $D_{a,n}$. 862 out of 982 EDs (88%) in several clinical EEG recordings were detected by this method together with 206 artifacts and 46 'pathological events' (events not relevant to epilepsy). A second stage of their system dropped 147 (68%) of the artifacts and 22 (2.5%) of the EDs. EDs were mapped on a 3D model of the patient's head. A series of maps and an animated map was used to display the evolution of potentials [Senhadji *et al.* 1994, Senhadji *et al.* 1995].

Clarencon *et al.* [1996] presented a real-time implementation of the CWT of 8-channel EEG using a *digital signal processor* (DSP). They applied Morlet's wavelet ($\psi(t) = e^{i\omega_0 t} \cdot e^{-t^2/2}$ with $\omega_0 = 5$, Fig. 2.4(a)) in dyadic scales and added three scales linearly between the dyadic ones (a total of 25 scales). Examples of spikes and seizures in animal EEG show broad energy distributions of a conical shape in the corresponding scalograms. The presence of an epileptic spike was characterized by an energy concentration between 10 and 15 Hz in the scalogram.

D'Attelis *et al.* [1997] applied a polynomial spline wavelet (Fig. 2.4(b)) for the DWT of EDs found in depth recordings of patients candidate to surgical treatment. For each scale they defined a detection threshold $D = \overline{x} + \sigma$ with the average $\overline{x}$ and the standard deviation σ of the energy of the wavelet coefficients in an 8 s period. Spikes appear in the scale level $j = -3$ which corresponds to a frequency range of 16 - 32 Hz. Waves following the spikes are represented by the scale level $j = -5$ with a frequency range of 4 - 8 Hz.

Sartoretto and Ermani [1999] used only a single scale of the DWT with the Daubechies-3 wavelet (16-32 Hz bandpass) to analyze 79 min of bipolar 8-channel EEG. The background activity was estimated from a 5 s epoch. Less than 5% of the recording time were marked by the system while 96% of EDs were detected.

2.3.10 Seizure Detection and Prediction

Gotman [1982] proposed a set of patient-specific features for seizure detection:

- the ratio of amplitude of current epoch to amplitude of the background,
- the average frequency of the EEG in the current epoch,
- the average frequency of the EEG in the background,
- the coefficient of variation of wave duration, and
- the ratio of amplitude of the current epoch to amplitude of next 8 s

and proposed a system which was demonstrated to have a false seizure detection rate of 3.3 per hour. In a subsequent study the feature classification was improved by the use of the *Mahalanobis distance* (Eq. 5.6) which reduced the false seizure detection rate to 1.3 per hour. Only 5% of seizures were missed in a data set of 2068 h of EEG recordings [Qu and Gotman 1993]. In further refinements of the algorithm a false detection rate of 0.21 per hour without missed detections was achieved. However, the system was patient-specific and a seizure sample was required to initialize the algorithm for each patient [Qu and Gotman 1995, Qu and Gotman 1997].

Gabor *et al.* [1996] used the DWT to identify seizures in long-term EEG recordings. The DWT (Daubechies-4 wavelet) of 8-s epochs was analyzed. Each scale was assigned a weight factor (0..1) reflecting the magnitude of contribution to a training set of seizures. In the evaluation process the weighted wavelet coefficients were inverted back into time-series. Eight FFTs of overlapping intervals of these signals were arranged in a frequency-time matrix and the 2D-FFT was calculated. The coefficients of the 2D spectrograms were fed to a modified SOFM. An average of 90% of seizures were detected without false detections. With increased detection sensitivity 98% of seizures were detected with 7.55 false detections per hour.

If seizures can be predicted, the generating mechanism may be understood and intervention may be possible. Iasemidis *et al.* [1996] showed that the *Lyapunov exponent* $L_{\max}$, (a measure of signal complexity in non-linear dynamical systems) of cortical grid recordings decreased several minutes before seizure onset. The exponent falls to a minimum and quickly rises to a maximum at seizure onset.

Geva and Kerem [1998] used the DWT to examine the EEG of rats with *hyperbaric oxygen* induced seizures. They used the first derivative of the Gaussian as a wavelet function and estimated the DWT according to the algorithm proposed by Mallat and Zhong [1992]. Several statistical parameters, such as *variance*, *skewness* and *kurtosis*, were computed from the coefficients of overlapping intervals and passed to an 'unsupervised optimal fuzzy clustering' algorithm. Sleep stages and seizures were represented by distinct cluster centres for each subject. In 16 of 24 rats studied a *pre-ictal state* could be identified 42 to 270 s before seizure onset.

Salant *et al.* [1998] investigated the coherence between two channels at the seizure frequency (3.5 Hz). They observed increased coherence corresponding to preictal synchronisation 1-6 s before seizure onset.

Lehnertz and Elger [1998] evaluated another measure used in non-linear dynamical systems, the *correlation dimension* D_2^{eff}, from cortical grid recordings of patients with pharmacoresistant epilepsy under pre-surgical evaluation. Seizures were preceded by a decreased time-resolved estimate for the correlation dimension in derivations recorded from within the primary epileptogenic area. A preseizure state could be identified 1-10 min prior to clinical seizure onset in 16 patients.

2.4 SUMMARY

Epilepsy is a neurological disorder that covers various epileptic syndromes. Most epileptic syndromes can be classified by the charateristic types of seizures observed.

In most epileptic patients, seizures produce major changes in the EEG. The interitcal EEG also shows characteristic epileptiform patterns of high diagnostic relevance in most patients.

Recording the interictal EEG provides an important diagnostic tool but the reading and interpretation of the EEG by an electroencephalographer (EEGer) is a highly skilled and time-consuming process. Many signal processing techniques have been used to automatically detect epileptiform activity in the EEG such as mimetic, parametric, syntactic, ANN-based and wavelet-based approaches. None of the systems documented in the literature have attained a performance equivalent to that of the human EEGer.

2.5 OBJECTIVE

The objective of the current work was to develop a system based on wavelet analysis of the multichannel EEG to determine whether the multiscale time-frequency attributes of this approach could lead to more accurate automatic detection of interictal epileptiform activity[2].

The system could function in two possible ways. Firstly, it could function as a screening tool for the neurophysiologist which reduces the number of pages reviewed. For this end a large sensitivity would be preferred while a small selectivity is acceptable. Secondly, the system could function as a long-monitoring system for which a small false detection rate would be required and lower sensitivity may be acceptable. The system should therefore be adjustable to cater for both options.

[2]Strictly speaking, a discharge can only be labelled 'epileptiform' if a clinical diagnosis of an epileptic syndrome has been made. In this work, *electrographic* epileptiform discharges will be referred to as definite EDs if they have been detected by at least two out of three clinical neurophysiologists in the EEG. The output of the system presented later in the thesis does not answer the question 'Does the patient suffer from epilepsy?' but rather 'Does the EEG show abnormalities of an epileptic character?'

Chapter 3

WAVELET-BASED SIGNAL ANALYSIS

3.1 INTRODUCTION

Wavelet-based signal analysis offers new ways to approach classical problems in signal processing by evaluating the time and the frequency domains simultaneously. This is desirable for non-stationary signals or signals with time-varying spectra.

Witkin [1983] investigated the behaviour of edges in signals through a new transform he called *scale space decomposition.* He could identify the amount of Gaussian blur needed to annihilate a local extremum in the derivative of an edge. Differentiating the Gaussian before convolving with the signal yields an optimal scalable edge detector which is now known as the DOG1 wavelet [Canny 1986].

Analysing a signal at several scales simultaneously amounts to an expansion of the signal in terms of several bandlimited frequency channels. This representation is somewhere between the time and frequency domains. Motivation to investigate such multifrequency channel decompositions came from psychophysiological research on the human visual system. Findings suggest that the retinal image is decomposed into several spatially-oriented frequency channels [Mallat 1989a].

Distinct types of wavelet transforms offer new perspectives for a wide range of signal processing applications such as compression, de-noising, time-frequency analysis, feature detection and feature extraction. Wavelet transforms are well suited to the representation of non-stationary physiological signals such as the EEG, the *electromyogram* (EMG) and *evoked potentials* (EPs). Similar representations cannot be achieved with the Fourier transform or *short-time Fourier transform* (STFT) because they are based on analyzing intervals of a fixed length.

There are two basic approaches to wavelet analysis: *discrete wavelet transform* (DWT) and the *continuous wavelet transform* (CWT). Both are presented in detail in the following sections.

The DWT uses a very efficient algorithm that repeatedly splits the spectrum with a filter pair (high-pass and low-pass). Thus the DWT will deliver a dyadic range of scales, i.e., each scale represents one octave of the spectrum. The inverse DWT delivers a perfect reconstruction of the original signal. Wavelets for the DWT are hard to construct because they must be orthogonal [Strang and Nguyen 1997]. However, a variety of orthogonal wavelets are available in the literature [Daubechies 1988, Cohen *et al.* 1992]. One is at liberty to choose wavelets which are near-symmetric or feature a predefined number of vanishing moments. One has to be careful in the interpretation of the coefficients delivered by the DWT since this transform is not translation-invariant.

For the CWT, one has more freedom in the choice of the wavelet function: real- or complex-valued wavelets may be applied. The relative modulation and the time-frequency bandwidth product can be chosen to match the application. The range of scales (overall bandwidth covered) and the progression of scales (e.g. linear, dyadic, or harmonic) must be selected. The interpretation of the CWT's output, the *scalogram*, is less straightforward than for the *spectrogram* resulting from a STFT. Both the frequency of an oscillation and the time a transient occurs can be read in a single scalogram. It is important to note the frequency bias that may be introduced by the normalization method of the wavelet transform [Grossmann *et al.* 1989, Daubechies 1990].

3.2 RELATIONSHIP TO THE SHORT-TIME FOURIER TRANSFORM

In the STFT a window function is applied to the signal to focus on a particular period in time. The STFT then analyzes all harmonics within that period by means of complex modulation and averaging. Frequencies are selected by changing the modulation. In contrast, in a wavelet transform a bandpass filter is *re-scaled* to focus on a particular frequency.

In a wavelet transform the signal is correlated with a *scaled* (or *dilated*) and *shifted* (or *translated*) version of a prototype pattern called the *wavelet function* (also *analyzing wavelet* or *mother wavelet*). Since no fixed window length is used, structures in a signal can be resolved and identified even if they are quite variable in scale.

For the implementation of the STFT a window function and a window length must be chosen. More choices have to be made if a wavelet transform is to be applied. Firstly, the the type of transform must be chosen: DWT or CWT. Then, depending on the type of transform, the wavelet function must be chosen. There is no general method to find the wavelet function best suited for an application. Interpretation of the transformed data is more complex than for the Fourier transform.

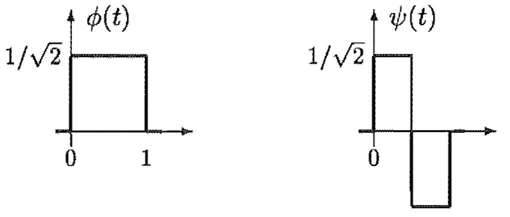

Figure 3.1 Scaling function $\phi(t)$ and corresponding wavelet function $\psi(t)$ of the Haar wavelet.

3.3 DISCRETE WAVELET TRANSFORM

The most basic orthogonal multiresolution expansion[1] was described by the mathematician Alfred Haar as early as 1910 [Haar 1910]. He found that a function $f(t)$ could be expanded into an infinite series of approximation coefficients c_k and detail coefficients $d_{j,k}$ according to

$$f(t) = \sum_{k=-\infty}^{\infty} c_k\, \phi(t-k) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_{j,k}\, \psi(2^j t - k) \tag{3.1}$$

where $k, j \in Z$. $\phi(t)$ and $\psi(t)$ are nowadays known as the scaling function and the corresponding wavelet function respectively (Fig. 3.1). This expansion can be regarded as a transform to a new basis formed by shifted and scaled instances of $\phi(t)$ and $\psi(t)$.

3.3.1 Multiresolution Decomposition

For digital signal processing a very short and efficient filter pair can be derived from the scaling and wavelet functions: $\phi(t)$ corresponds to the low-pass filter H_0 and $\psi(t)$ to the high-pass filter H_1. Both filters are applied to a signal and the resulting signals are subsequently downsampled by a factor of two (by removing every other sample). The signal is effectively split up into an approximation signal and a detail signal (Fig. 3.2). It is important to note that the overall number of samples is maintained in the decom-

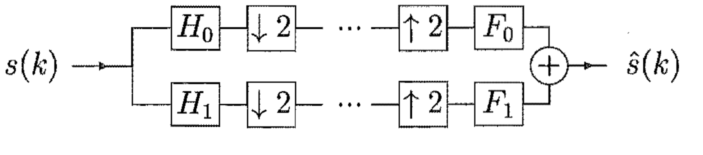

Figure 3.2 Two-channel filter bank. The input signal is filtered with a pair of decomposition filters (H_0 and H_1) and subsequently downsampled. The result is an approximation and a detail signal. Upsampling and application of reconstruction filters F_0 and F_1 returns the input signal.

position. The original signal can be recovered by an upsampling step (inserting zeros between samples) and the application of the corresponding reconstruction filter pair F_0 and F_1, which are time inverses of H_0 and H_1, respectively. Application of the time

[1] An orthogonal expansion breaks a function down into a number of uncorrelated components. In a multiresolution expansion those components vary in length.

inverses is important to restore the phase of the original signal. Another important property of the filter pairs is that both convolutions $H_0 * F_0$ and $H_1 * F_1$ produce a delayed unit response after downsampling.

When the decomposition step is repeated for the approximation signal several times a single approximation signal and a series of detail signals are acquired (Fig. 3.3). This procedure is known either as the DWT or *multiresolution decomposition* (MRD).

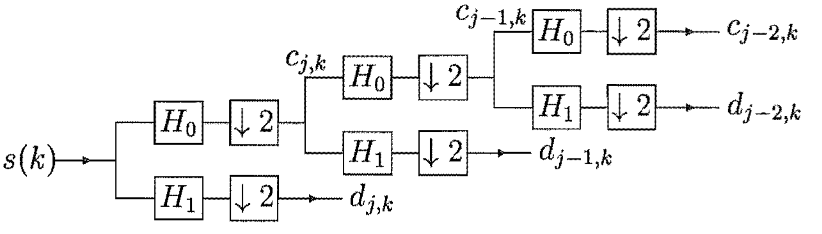

Figure 3.3 Multiresolution Decomposition structure. The filtering and downsampling steps are repeated several times for the approximation signal.

The first approximation signal $c_{j,k}$ is similar to a signal sampled at half the original sampling rate[2]. The second approximation $c_{j-1,k}$ corresponds to a quarter of the original sampling rate, and so on. High frequency details in the approximation signal are gradually lost as we progress through the levels of the decomposition. In fact, these details are precisely retained in the detail signals $d_{j,k}$, so the original signal can be reconstructed from the wavelet coefficients. The inverse transform is performed in a series of steps consisting of upsampling, filtering with F_0 and F_1 and addition of the detail signals.

In the DWT only dyadic scales $a = 2^j$, $j \in N$, and translations $b = n2^j$, $n, j \in N$, are computed (refer to Eq. 3.3 for definition of a and b). For orthogonal wavelet functions, this subset of points in the time-scale plane forms an orthogonal basis. Thus, the coefficients obtained by the DWT are uncorrelated with each other and the transform can be inverted.

The DWT is a highly efficient algorithm. The order of computational complexity is only $O(N)$, i.e., the number of operations rises only linearly with the number of samples in the signal.

3.3.2 Wavelets for the DWT

An *orthogonal wavelet* must be chosen for the DWT to ensure its invertability. A wavelet function $\psi(t)$ is orthogonal if

$$\int_{-\infty}^{\infty} \psi_{jk}(t)\, \psi_{JK}(t)\, dt = \delta(j-J)\delta(k-K) \tag{3.2}$$

[2]Note that the sample index k does not refer to the same time instance for the input and output signals.

with $\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$ and $j, k \in Z$. Orthogonality ensures a complete partitioning of the frequency domain and thus a unique and complete representation of the signal. Orthogonality also imposes strong constraints on the wavelet function and the corresponding discrete filters. A complicated mathematical framework is needed to construct orthogonal wavelets and wavelet filters [Strang and Nguyen 1997].

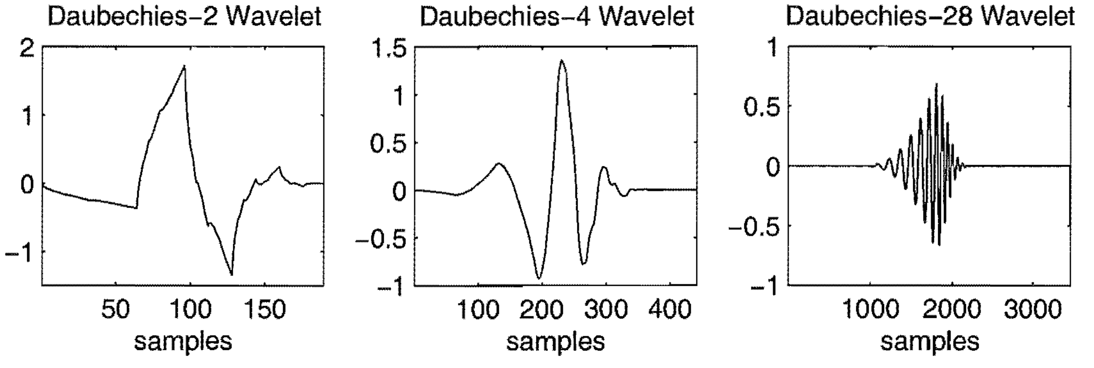

Figure 3.4 Level-5 reconstructions of Daubechies wavelets filters of order $n= 2$, 4, and 28.

3.3.3 Daubechies Wavelet Family

Daubechies developed a family of wavelets that combine optimal band-separation (or maximal 'flatness' of the filter characteristics) with short filter lengths [Daubechies 1988]. These wavelets are commonly referred to as Daubechies-n wavelets, where n indicates the order of the wavelet (Fig. 3.4). The filters corresponding to the Daubechies wavelet of order n feature $2n$ non-zero coefficients. The Haar wavelet is equivalent to the Daubechies wavelet of order 1, or Daubechies-1 wavelet. The considerable overlap between the low-pass transfer function and the high-pass transfer function for the Haar

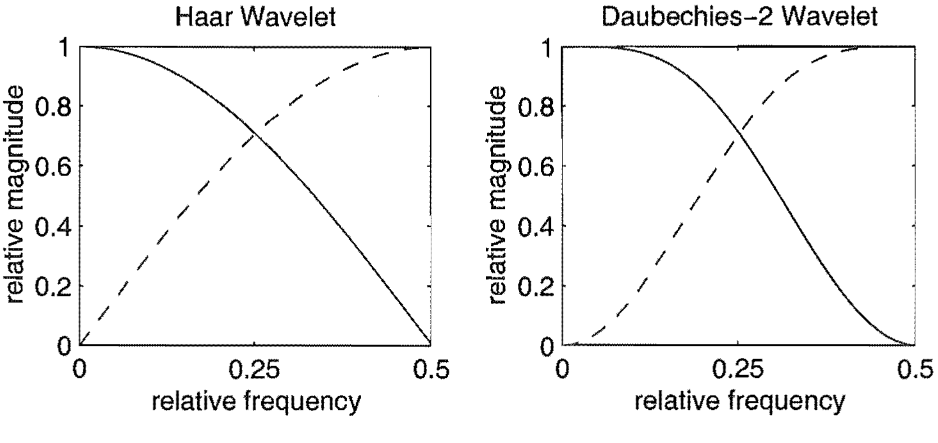

Figure 3.5 Frequency characteristics of the Haar wavelet filter pair and the Daubechies wavelet filter pair of order n= 2 (solid lines: low-pass, dashed lines: high-pass). The quality of the band separation increases with the order of the wavelet.

wavelet (Fig. 3.5 left) is gradually overcome by the higher order Daubechies wavelets

(Fig. 3.5 right and Fig. 3.6). Note in Fig. 3.4 that the number of oscillations increases with the order of the wavelets. No analytic expressions exist for the wavelet functions

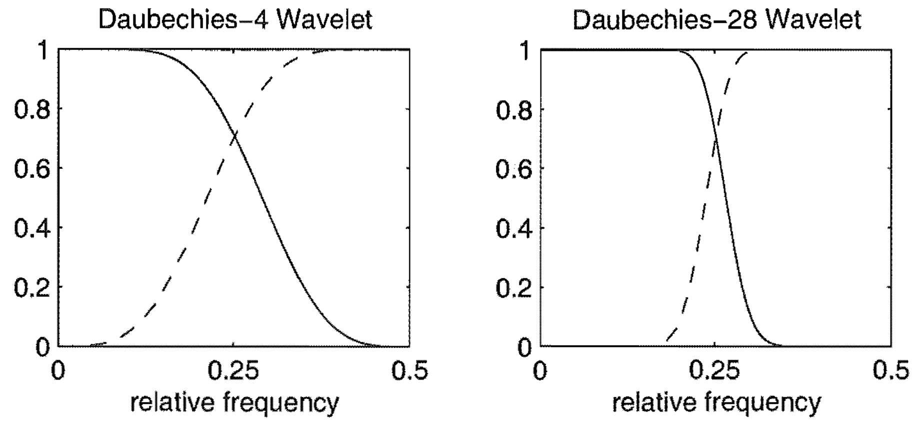

Figure 3.6 Frequency characteristics of the Daubechies wavelet filter pair of orders $n = 4$ and $n = 28$ (solid lines: low-pass, dashed lines: high-pass).

of the Daubechies family. However, the filter pairs can be determined exactly for any order n and the wavelet functions can be recursively approximated from the filters [Strang and Nguyen 1997].

3.3.4 Aliasing in the DWT

Although the frequency characteristics of the decomposition filter pairs are smooth, the downsampling step of the MRD causes aliasing in the overall decomposition (Fig. 3.7). For the Daubechies-2 wavelet the aliasing corresponds to the high-frequency detail

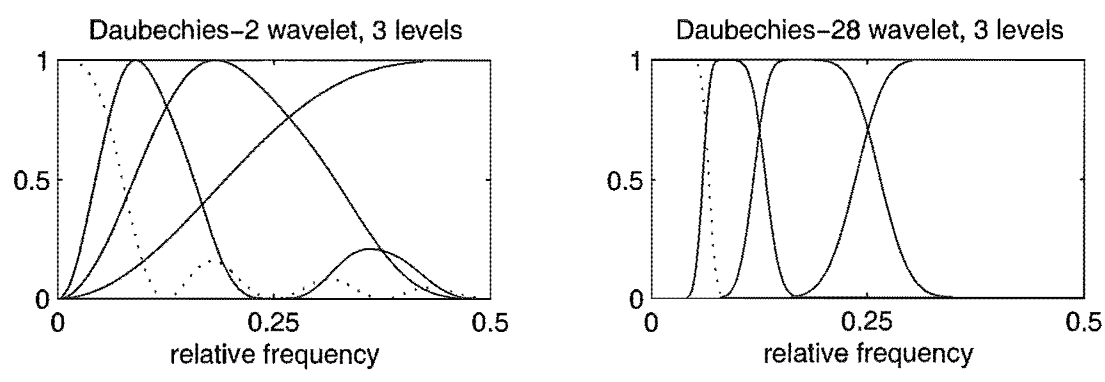

Figure 3.7 Transfer function characteristics of 3-level decomposition with the Daubechies wavelet filter pairs of order $n = 2$ and $n = 28$ (dotted lines: approximations, solid lines: details).

structure of the waveshape in Fig. 3.5. The aliasing is well attenuated for the higher-order Daubechies wavelets. Strang and Nguyen [1997] point out that a *binomial* or *maxflat filter* is used in the construction of the Daubechies wavelet filters. Thus their frequency responses are *maximally flat* at relative frequency 0.5 or, in other words, the aliasing is minimal for the Daubechies filters with respect to other filters of the same length.

3.3.5 Symmetric, Biorthogonal and More Wavelets

Except for the trivial case of the Haar wavelet, orthogonal wavelets cannot be symmetric [Strang and Nguyen 1997]. Near-symmetric orthogonal wavelets called ***coiflets*** were proposed by Daubechies [1992].

For orthogonal wavelets, decomposition and reconstruction filters are time-inverse versions of each other: $H_0(z) = F_0(-z)$ and $H_1(z) = F_1(-z)$. If these filter pairs are modified in a way such that the result of the convolution $H_0 * F_0$ is unchanged, a biorthogonal filter pair is obtained. The corresponding ***biorthogonal wavelets*** are less constrained but nevertheless give perfect reconstruction. Biorthogonal wavelets can be constructed in a fully symmetric fashion [Cohen *et al.* 1992].

Recently, a general and straightforward procedure to construct all orthonormal filter pairs of even length was presented by Sherlock and Monro [1998]. Their procedure converts a set of N angles into a filter pair of length $2N$. Thus, they actually cover all degrees of freedom for filters usable with the DWT.

3.3.6 Response of the DWT to Non-Dyadic Signal Translations

The DWT leads to a discontinuous subdivision of the time domain. So a translation of the signal by a prime number of samples $\Delta b > 2$ causes a change in all coefficients of the DWT. A detection procedure based on the DWT can only fully respond to some target pattern if that pattern occurs at some dyadic translation coinciding with the temporal alignment of the DWT [Thakor and Sherman 1995].

3.3.7 Wavelet Packets

The MRD splits only the approximations $c_{j,k}$ at each level with a filter pair. When both $c_{j,k}$ and $d_{j,k}$ are split the result is a binary tree structure. In a ***wavelet packet***

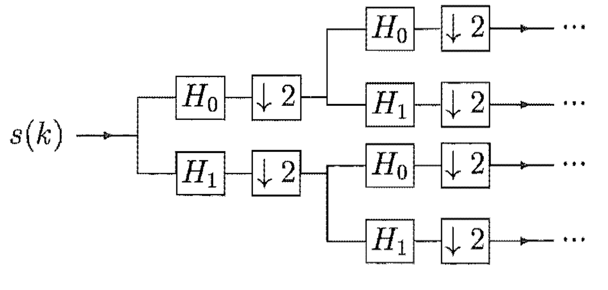

Figure 3.8 Full decomposition of approximation and detail signals results in a binary tree structure. The tree is pruned to optimize the decomposition.

decomposition each node is either retained or removed according to some criterion (e.g. entropy). Thus, an adapted basis is formed which can be useful in compression and de-noising applications.

3.4 CONTINUOUS WAVELET TRANSFORM

The DWT can only give a crude picture of the time-frequency content of a signal, e.g. all twelve musical tones in an octave will be covered by a single scale. A more detailed picture is obtained with the CWT, where the scale parameter a can be varied in minute steps. This approach is much more closely related to the original concept of a continuous *scale space* introduced by Witkin [1983].

The CWT of a continuous input signal $s(t)$, $t \in R$, and a complex wavelet $\psi(t)$, $t \in R$, can be defined as

$$\mathrm{CWT}(b, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t)\, \psi^* \left(\frac{t-b}{a} \right) dt \tag{3.3}$$

where a denotes the scale parameter and b the time location ($a, b \in R$, $a > 0$). For a discretely sampled (bandlimited) input signal $s(k)$, $k \in Z$, we can compute the CWT according to

$$\mathrm{CWT}(k, a) = \frac{1}{\sqrt{a}} \sum_{l \in Z} s(l)\, \psi_a^*(l-k) \tag{3.4}$$

where

$$\psi_a(k) = \frac{1}{\sqrt{a}} \int \mathrm{sinc}(t-k) \psi \left(\frac{t}{a} \right) dt \tag{3.5}$$

is the bandlimited instance of the wavelet ψ at scale $a \in R$ [Ho and Chan 1999].

The inverse transform of the CWT exists in theory, but its implementation is numerically unstable. Importantly, the CWT is translation-invariant, which makes it the preferred choice for applications like transient detection and time-frequency analysis. However, unlike the DWT, the CWT is a redundant representation of the signal. Since there is no downsampling step involved in the CWT, the transform does not produce aliasing unless it is a feature of the wavelet used. The time-scale representation resulting from the CWT is a *scalogram*. It is acquired by calculating the convolution (Eq. 3.4) for a range of scales a.

3.4.1 Wavelet filters for the CWT

There are fewer constraints for the construction of a wavelet function for the CWT than for the DWT. Grossmann *et al.* [1989] list the basic requirements for functions to be employed as wavelets and give examples of wavelet transforms. A wavelet is 'admissible' for use with the CWT if it has zero mean and is localized in both time and frequency domains.

Several important wavelet functions for the CWT are generated from the Gaussian by temporal differentiation or modulation (both real and complex). Others, including the binomial kernel wavelet family, introduced by the author [Goelz *et al.* 2000a], employ other envelopes to achieve time localization. Wavelet filters are derived directly by sampling the explicit expressions of the wavelet functions (Eq. 3.5).

Important features of the wavelets are their time-frequency bandwidth products $\sigma_t^2 \cdot \sigma_f^2 \cdot (4\pi)^2$ and their relative bandwidths Q. These are defined in subsection 3.4.6 and listed for several of the wavelet filters considered.

3.4.2 Derivatives of the Gaussian

The first temporal derivative of the Gaussian (*DOG1*)

$$\psi(t) = -t \cdot e^{-t^2/2} \tag{3.6}$$

is antisymmetric and leads to optimal edge detection filters [Canny 1986]. The second temporal derivative (*DOG2*)

$$\psi(t) = (1 - t^2)e^{-t^2/2} \tag{3.7}$$

is called the *Mexican hat wavelet* or *Marr wavelet*. It is important for transient detection applications (Fig. 3.9).

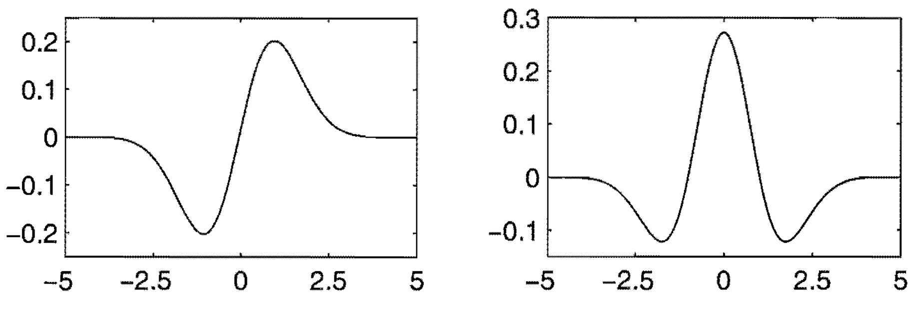

Figure 3.9 DOG1 (Gaussian pulse) and DOG2 (Mexican hat) wavelet filters.

3.4.3 Modulation of the Gaussian

The modulated Gaussian or *Gabor-function*

$$\psi(t) = \pi^{-1/4} e^{i\omega_0 t} \cdot e^{-t^2/2} \tag{3.8}$$

is called *Morlet wavelet*. It features optimal time-frequency resolution [Grossmann *et al.* 1989, Daubechies 1990]. The modulation can be adjusted through $\omega_0 \geq \pi\sqrt{2/\ln 2}$, which imposes a lower limit on the relative bandwidth of the wavelet filter (Fig. 3.10). Note that the Gaussian does not have strictly finite support, so an arbitrary truncation

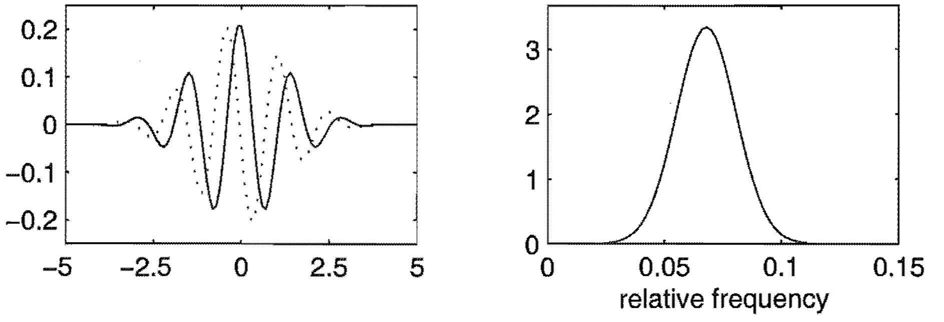

Figure 3.10 Left: Morlet wavelet filter generated from complex modulated Gaussian (modulation $\omega_0 = \pi\sqrt{2/\ln 2}$). Right: detail of corresponding frequency response.

has to be done in implementations. This truncation is, however, equivalent to the application of a rectangular window and the corresponding aliasing in the frequency domain.

Another complication is to find the smallest admissible scale (or highest frequency) at which the wavelet is represented after sampling (Eq. 3.5). In other words, it is important to make sure the scaled instance of the wavelet fits into the bandlimited discrete domain.

3.4.4 Modulation of Other Window Functions

Window functions other than the Gaussian can be used to build wavelet filters. A modulated Hanning window for instance

$$\psi_k(t) = \begin{cases} (1+\cos(2\pi t)) \cdot e^{i2\pi kt} & \text{if } |t| < \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

with integer $k \neq \{-1, 0, 1\}$, leads to a transient-like complex wavelet filter (Fig. 3.11). For $k = 2$ the relative bandwidth (Tab. 3.1) is smaller than for the Morlet wavelet. This wavelet is defined in a limited temporal interval or, in other words, it has *compact support* in contrast to the Morlet wavelet [Senhadji *et al.* 1995].

The smallest possible relative bandwidth of a modulated window can be realized with the 'psi_1' wavelet devised by the author for the decomposition of *event related potentials* (ERPs) [Gölz 1995, Gölz *et al.* 1997]. The psi_k wavelet filters are generated

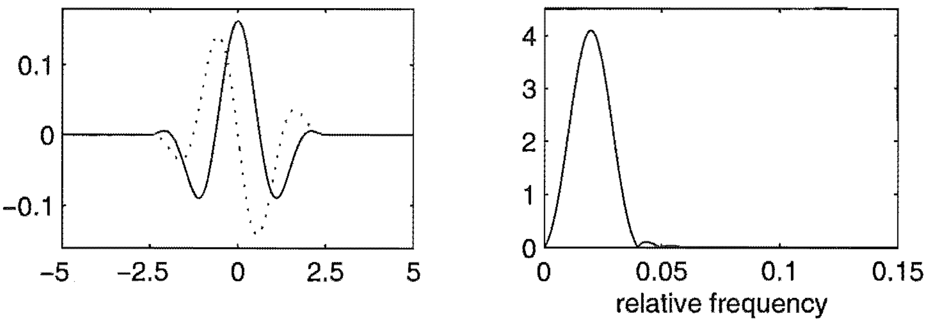

Figure 3.11 Left: wavelet filter generated from complex modulated Hanning window (modulation $k = 2$). Right: detail of corresponding frequency response [Senhadji *et al.* 1995].

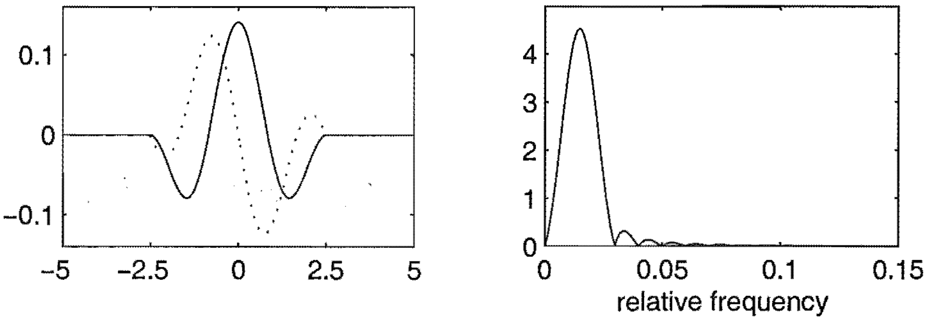

Figure 3.12 Left: psi_k wavelet filter generated from a complex modulated sine halfwave (modulation $k = 1$). Right: detail of corresponding frequency response.

from complex modulated sine halfwave

$$\psi_k(t) = \begin{cases} \frac{1}{2}(e^{i2\pi kt} + e^{i2\pi(k+1)t}) & \text{if } |t| < \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{3.10}$$

with $k = 1, 2...$ (Fig. 3.12). For the smallest modulation, $k = 1$, there are 1.5 oscillation cycles on the compact support of the wavelet. The real part correlates well with the Mexican hat ($r > 0.99$).

For integer scales, the sine halfwave and the Hanning window can be regarded as discrete Fourier transforms of zero-padded binomial sequences

$$B_n(m) = \binom{n}{m} \tag{3.11}$$

of orders $n = 1$ and $n = 2$ respectively. Thus, we can regard the corresponding wavelets as members of a more general *binomial kernel wavelet family* [Goelz *et al.* 2000a]

$$\psi(t, m_0, n) = \begin{cases} \sum_{m=0}^{n} \binom{n}{m} e^{j2\pi(m+m_0)t} & , \text{if } |t| < \frac{1}{2} \\ 0 & , \text{otherwise} \end{cases} \tag{3.12}$$

with modulation $m_0 \geq 1$. This wavelet family features a small relative bandwidth for small n and approaches the optimal time-frequency bandwidth product for large n (see section 3.4.6). It is important to note that the modulation limit $m_0 \geq 1$ creates a

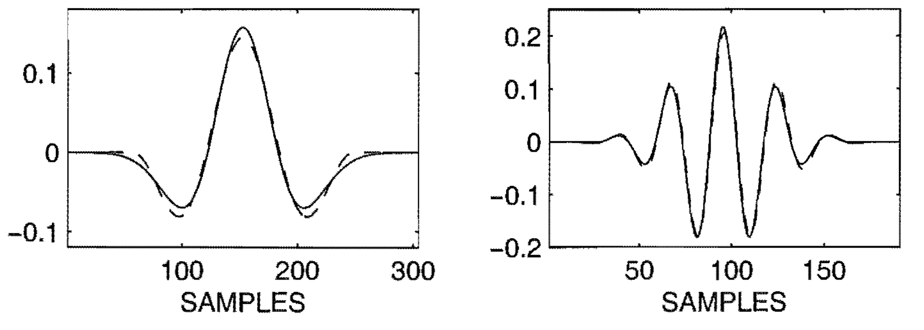

Figure 3.13 Left: correspondence of binomial kernel wavelet of order $n = 1$ (modulation $m_0 = 1$) and Mexican hat wavelet: $r = .993$. Right: correspondence of binomial kernel wavelet of order $n = 5$ (modulation $m_0 = 3$) and Morlet wavelet: $r = .993$.

DC-free bandpass filter. Furthermore, for a given order n and modulation m_0 there is a well defined lower limit for the scales available in the CWT: the length of the filter N must be larger than $2(m_0 + n)$. This confines the binomial kernel to the positive part of the spectrum.

3.4.5 Wavelets based on B-Splines

Repeated convolution of the box function

$$\beta^0(t) = \begin{cases} 1 & , \text{if } -\frac{1}{2} \leq t < \frac{1}{2} \\ 0 & , \text{otherwise} \end{cases} \tag{3.13}$$

creates a B-spline of degree n

$$\beta^n(t) = \beta^0 * \beta^{n-1}(t) \tag{3.14}$$

which approximates the Gaussian for large n (Fig. 3.14). Likewise, a modulated B-

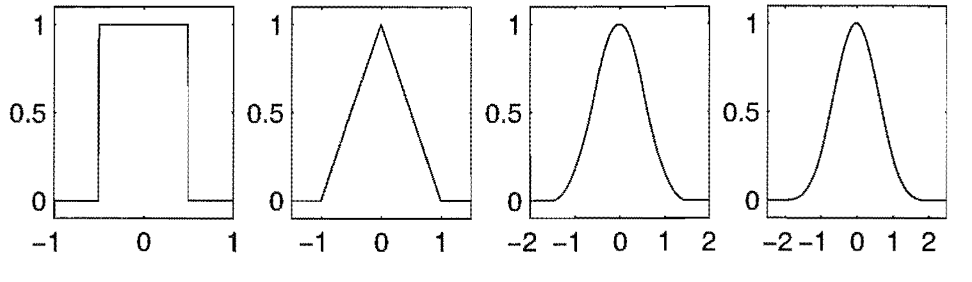

Figure 3.14 B-Splines of order n= 0, 1, 2 and 3. B-Splines have been normalized to a maximum of one.

spline

$$\psi(t) = \beta^n(t)\, e^{j\omega t} \tag{3.15}$$

can approximate the Morlet wavelet as closely as desired. The minimum modulation can be achieved with $\omega = 2\pi$. The corresponding filters are derived directly from the discrete B-splines and can be computed by a fast and scale-independent algorithm for integer scales [Unser *et al.* 1993a, Unser *et al.* 1993b, Unser 1994].

B-Splines also provide an excellent means of interpolating a discretely sampled signal, which involves transformation of the signal to the B-Spline domain as an initial step. In the B-Spline domain the DOG1 and Mexican Hat wavelets can be implemented as fast filters [Unser *et al.* 1994b, Schiff *et al.* 1994a, Unser *et al.* 1994a, Unser 1999].

3.4.6 Time-Frequency Resolution and Relative Bandwidth

According to Heisenberg's uncertainty principle, there is a lower bound for the simultaneous localization of an event with duration σ_t in the time domain and bandwidth σ_f in the frequency domain:

$$\sigma_t^2 \cdot \sigma_f^2 \geq \frac{1}{(4\pi)^2} \; . \tag{3.16}$$

The equality is only achieved with Gaussian waveforms. The *time-frequency bandwidth product* $(4\pi\sigma_t\sigma_f)^2$ represents the intrinsic blur of the scalogram: a single scaled and translated instance of the wavelet does not appear as a single dot in the scalogram but leads to a two dimensional peak [Unser *et al.* 1992, Unser 1994]. The area covered by this peak is minimal for the Morlet wavelet; thus it features minimal blur and optimal time-frequency resolution [Daubechies 1990]. For a discrete wavelet filter $w(k)$ the time duration can be estimated by the temporal standard deviation

$$\sigma_t^2 = \frac{\sum_k |w(k)|^2\,(k - \overline{w})^2}{f_s^2 \sum_k |w(k)|^2} \tag{3.17}$$

with the first moment

$$\overline{w} = \sum_k k \cdot |w(k)|^2 / \sum_k |w(k)|^2 \tag{3.18}$$

and sampling frequency f_s. The bandwidth can be estimated by the standard deviation of the spectrum

$$\sigma_f^2 = \frac{f_s^2 \sum_l |W(l)|^2\,(l - \overline{W})^2}{a^2 \sum_l |W(l)|^2} \tag{3.19}$$

with $W(l)$ being the *discrete Fourier transform* of $w(k)$ with mean frequency

$$\overline{W} = \sum_l l \cdot |W(l)|^2 / \sum_l |W(l)|^2 \ . \tag{3.20}$$

To compare the localization properties of the wavelet filters, the *relative bandwidth*

$$Q = \frac{\overline{W}}{2\sigma_f} \tag{3.21}$$

is calculated for the wavelet filters [Frisch and Messer 1992]. Since σ_f is fixed for a given window function a wavelet with a small relative bandwidth, Q can be obtained by using the smallest possible modulation (ω_0 in Eq. 3.8, k in Eq. 5.10 and Eq. 3.10) to minimize the mean frequency $\overline{W}$. The relative bandwidth Q relates to the effective number of oscillation cycles within the window function. A wavelet filter with a small Q is optimized for time localization of transients, while a large Q is optimal for frequency localization. The time-frequency bandwidth products and the relative bandwidths are listed in Table 3.1 and shown in Fig. 3.15.

Table 3.1 Time-Frequency Bandwidth product $(4\pi\sigma_t\sigma_f)^2$ and relative bandwidth Q of a selection of wavelet filters for the CWT.

Wavelet	parameter or order	$(4\pi\sigma_t\sigma_f)^2$	Q
DOG1		1.361	1.185
DOG2/Mexican hat		1.104	1.547
Morlet	$\omega_0 \geq \pi\sqrt{2/\ln 2}$	1.000	≥ 3.536
Binomial Kernel	$n = 1$	1.123	1.680
	$n = 2$	1.056	1.749
	$n = 3$	1.025	1.866
	$n = 4$	1.013	1.985
	$n = 5$	1.008	2.100
Modulated B-Spline	$n = 1$	1.209	1.741
	$n = 2$	1.015	2.348
	$n = 3$	1.004	2.660
	$n = 4$	1.003	2.957
	$n = 5$	1.002	3.222

If a complex wavelet filter needs to be implemented for a given relative bandwidth, the Morlet wavelet covers the range of all relative bandwidths above 3.5 with the optimal time-frequency bandwidth product. In the range $2.3 \leq Q \leq 3.5$ the modulated B-spline wavelets provide the best time-frequency bandwidth product, but for $Q < 2.3$ the binomial kernel wavelets are the best choice.

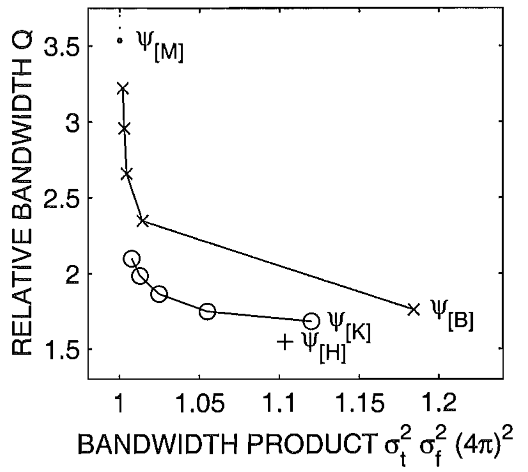

Figure 3.15 Relative bandwidth Q over time-frequency bandwidth product $(4\pi\sigma_t\sigma_f)^2$ of wavelet filters: Mexican hat [H], Morlet [M], modulated B-spline [B] and binomial kernel [K].

3.5 SCALE AND FREQUENCY COVERAGE

Scaling schemes other than linear or dyadic are commonly used with the CWT. Unlike the DWT, which covers the full bandwidth exactly, the CWT focuses on a limited frequency range only. Grossmann *et al.* [1989] suggested a harmonic (or geometric) progression of scales

$$a_n = 2^{n/s} \tag{3.22}$$

with s scales per octave. The set of scales applied a_n together with the relative bandwidth Q of the wavelet define the overall coverage of the spectrum.

3.5.1 Normalization of Wavelet Coefficients

The normalization factor $1/\sqrt{a}$ in the CWT (Eq. 3.4) ensures that the power of the wavelet filter will remain constant for all scales. This way, for the DWT, the basis is normalized and energy is preserved under the transform. However, if frequencies or amplitudes need to be measured the normalization $1/a$ is appropriate. Thus most authors define the CWT with the normalization which suits their applications. Grossmann *et al.* [1989] use both methods and say 'it is sometimes convenient' to apply one or the other normalization method. It is important to realize that only $1/a$ normalization can result in a correct reading of amplitudes and the a peak in the scalogram will be shifted (biased) if the other method is used. This is illustrated in detail in the following chapter.

3.6 MATCHING PURSUIT

The analyzing wavelet can either be chosen to have a small Q (few oscillation cycles) for transient detection or a large Q (many oscillation cycles) to improve the frequency discrimination in the CWT. If transients occur in a signal and a good frequency resolution is required at the same time, it would be advantageous to vary both scale and modulation of the analyzing wavelet[3].

Mallat and Zhang [1993] proposed a new method called *matching pursuit* which is not limited to a single analyzing wavelet. It uses, instead, a dictionary of waveforms, all of which are normalized to unit energy. The signal is filtered with all waveforms and the sample with the largest magnitude among all responses is chosen. An instance of the corresponding waveform is translated and scaled appropriately and is called a time-frequency atom of the signal. The atom is subtracted from the signal and added to a reconstruction signal. The procedure is repeated until the difference between original and reconstruction becomes sufficiently small.

Matching pursuit is the best way to see both "the forest and the trees" in a signal but it is very computationally expensive. It is an order of magnitude slower than the CWT, since all filters have to be applied to the residual signal at each iteration.

3.7 SUMMARY

With the aid of wavelet transforms a signal can be analyzed in the time domain and frequency domain simultaneously. Thus, non-stationary features such as transients and short bursts of oscillatory activity are better represented than with more conventional Fourier methods.

There are two major types of transforms:

- The *discrete wavelet transform* (DWT) is fast, invertible, not translation-invariant, and delivers one detail signal per octave and requires an orthogonal wavelet. It is well suited for applications like compression and de-noising.

- The *continuous wavelet transform* (CWT) is slower than the DWT, not invertible in practice, translation-invariant and allows for the application of complex wavelets which are optimal in terms of their time-frequency localization. It is well suited for detection and time-frequency analysis.

For a particular application, wavelets can be chosen from a large variety of waveforms. Fast algorithms for the convolution within the CWT are available for the most important wavelets.

[3]One is tempted to define a three-dimensional time-scale-modulation space here.

Matching pursuit decomposes a signal into a set of time-frequency atoms, which are iteratively selected from a dictionary of waveforms. Thus, transients and oscillations are well represented in a single transform but at a high computational cost.

Chapter 4

APPLICATIONS OF WAVELET-BASED METHODS

4.1 INTRODUCTION

It is important to choose an appropriate wavelet-based signal processing method for a specific application and to identify an appropriate analyzing wavelet. The DWT is optimal for compression applications and applications that require a fast inverse transform. However, for detection and feature extraction applications the redundant CWT is more suitable. Since our main interest lies in the detection of a transient biological waveform, such a waveform is simulated in this chapter so that the results of the DWT and CWT can be studied.

With the CWT the frequency of short oscillations can be estimated and it is possible to draw a line between transient and oscillatory events. These issues are crucial for applications to physiological signals. Wavelet analysis can be used to detect epileptiform spikes and sharp waves in the EEG but it also gives clues to identify and discriminate short alpha bursts.

The CWT with a complex wavelet is a relatively new tool for time-frequency analysis. How does it compare with conventional time-frequency distributions such as *Wigner-Ville* or *Choi-Williams*? A comparison is presented late in the chapter.

4.2 TRANSIENT DETECTION

Canny [1986] describes several criteria for an edge-detection operator he wished to optimize. An edge embedded in white Gaussian noise is best detected with a difference-of-boxes operator, which is equivalent to the Haar wavelet. He found a tradeoff between the detection probability and the quality of the localization of an edge by scaling the operator. With larger scales, the noise is better attenuated (wider integration) while the width of the detection peak grows leading to poorer localization. In addition to good detection and good localization he proposed a third criterion: there should only be a single local maximum in response to a single edge.

Since the Haar wavelet is sensitive to high frequencies, there are numerous maxima in the vicinity of an actual edge caused by noise. He numerically optimized the edge detection operator for a given *signal to noise ratio* and found an optimal detector that is close to the first derivative of the Gaussian (DOG1).

The detection of epileptiform spikes in the EEG is complicated by two things. Firstly, there is no single prototype pattern for an epileptiform discharge that could be used to optimize an operator and, even if it did exist, it could never be measured without noise. Secondly, the spectral composition of the background can change rapidly over time [Gotman 1985]. An optimal operator, such as the one proposed by Canny, is based on a known (or at least fixed) noise spectrum [Canny 1986].

Frisch and Messer [1992] examined the performance of several detectors with respect to a 'not-perfectly-known signal' in Gaussian white noise. A downsampled variant of the CWT (Morlet wavelet, 4 scales per octave, an 'almost orthogonal' basis) was used as signal representation [Daubechies 1990]. They concluded that in cases where prior knowledge about the transient time-frequency bandwidth product and relative bandwidth exists, the 'wavelet-representation-based detector' performs better than any other [Frisch and Messer 1992].

4.2.1 Construction of a Test Signal

To study the detection and representation properties of the various transforms, a test signal with transient and background features was constructed. This test signal was subsequently analyzed via the Fourier transform, the STFT, the DWT, and the CWT. Four transient waveforms were added to coloured noise (Fig. 4.1). Transients A and B were derived from the same waveform but feature slightly different amplitudes. The coloured noise (RMS= $6.9\,\mu V$) was derived from low-pass filtered white noise (low-pass of order 25, cut-off frequency 18.2 Hz). Note the variety of small transients in the coloured noise. The SNRs of the individual transients were 11.2 dB for spike A, 12.8 dB for spike B, 1.1 dB for wave C and 6.8 dB for transient D. Transients C and D were generated by sine-modulated Hanning windows (modulation $k = 1$, width $a = 105$ samples and modulation $k = 2$, width $a = 35$ samples, respectively). The SNR of all transients (148 non-zero samples) with respect to the noise was 4.88 dB, the overall SNR of the signal containing the transients (900 samples) was -2.96 dB.

4.2.2 Fourier Transform

The *discrete Fourier transform* (DFT) of the test signal is shown in (Fig. 4.2. While the individual transients C and D (Fig. 4.2 a) show a rather tight frequency localization in their DFT, there is a fairly broad distribution for spikes A and B. Adding the four transients has a strong effect on the resulting DFT (Fig. 4.2 b): the phase distributions

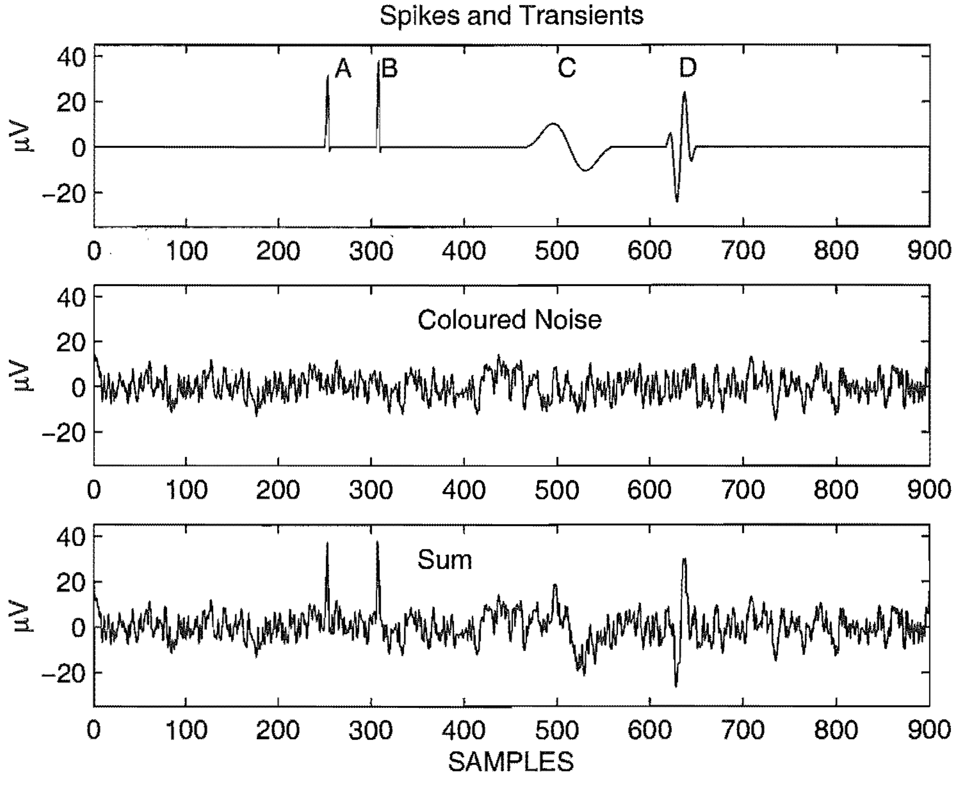

Figure 4.1 Four transients 'A' to 'D' (top row), coloured noise (middle row) and test signal (bottom row) composed of the sum of both. The simulated waveforms represent two monophasic spikes, one biphasic spike and a slow artifact. The test signal is equivalent to 4.5 s of EEG sampled at 200 Hz.

of all spectra interfere, so the well-defined spectral peaks of C and D appear as if they were distorted by noise. The spectra of spikes A and B interfere in a regular fashion

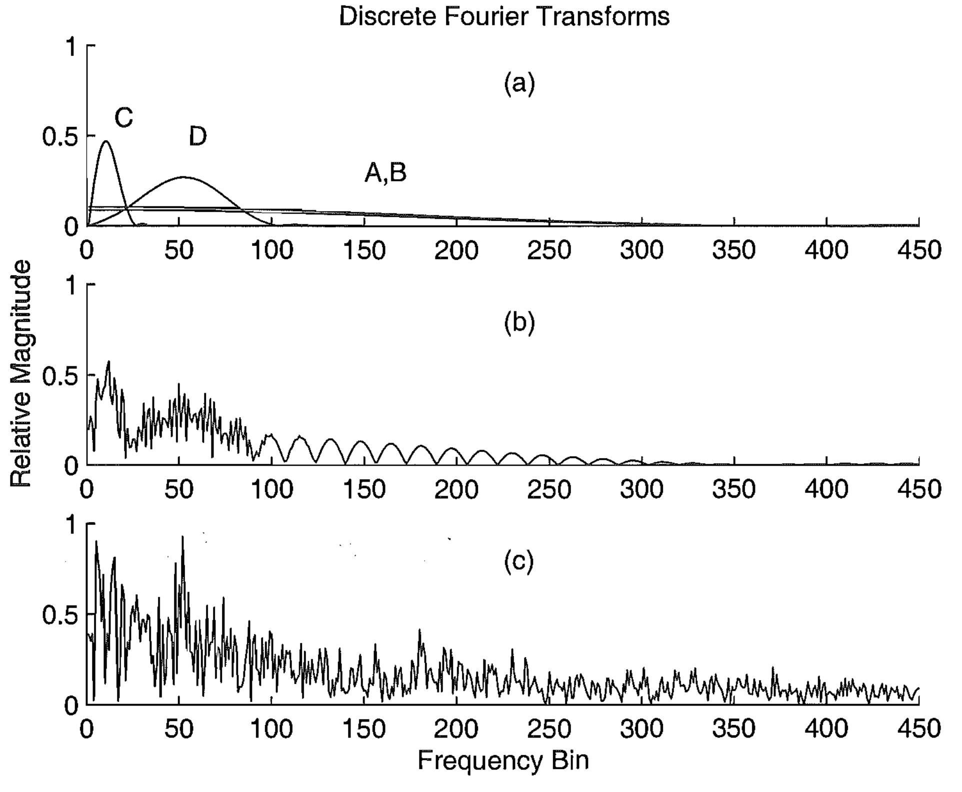

Figure 4.2 Discrete Fourier Transforms of test signal: (a) separate components A, B, C and D, (b) sum of components, (c) sum of components and noise.

between frequency bins 100 and 300. After addition of the noise the spectrum of the transients is mainly obscured (Fig. 4.2 c).

4.2.3 Short-Time Fourier Transform

The STFT was calculated using a window width of 64 samples (320 ms), so the test signal was split into 27 segments (Fig. 4.3). There was a 50% overlap between segments. Transient D is clearly visible in the STFT, while the other transients generate weak responses only. Transients A and B are spread out over a wide range of frequencies like in the overall Fourier transform. Transient C is too wide to fit into any of the segments so it is spread over several consecutive segments. When the window width is reduced to 32 samples (55 segments, again with a 50% overlap) a stronger response for the short transients A and B is obtained (Fig. 4.4). Now even the fourth transient does not fit into a single window. Thus, it is not possible to detect all four transients with a single window size appropriately.

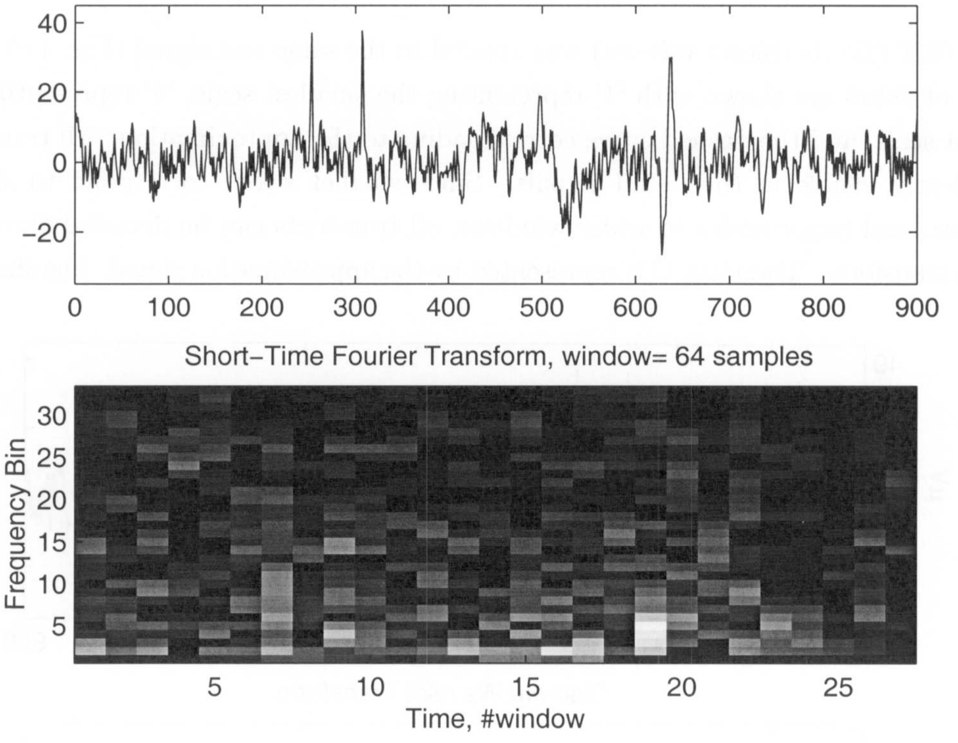

Figure 4.3 Test signal (top row) and modulus of the Short-Time Fourier Transform with window width 64 samples.

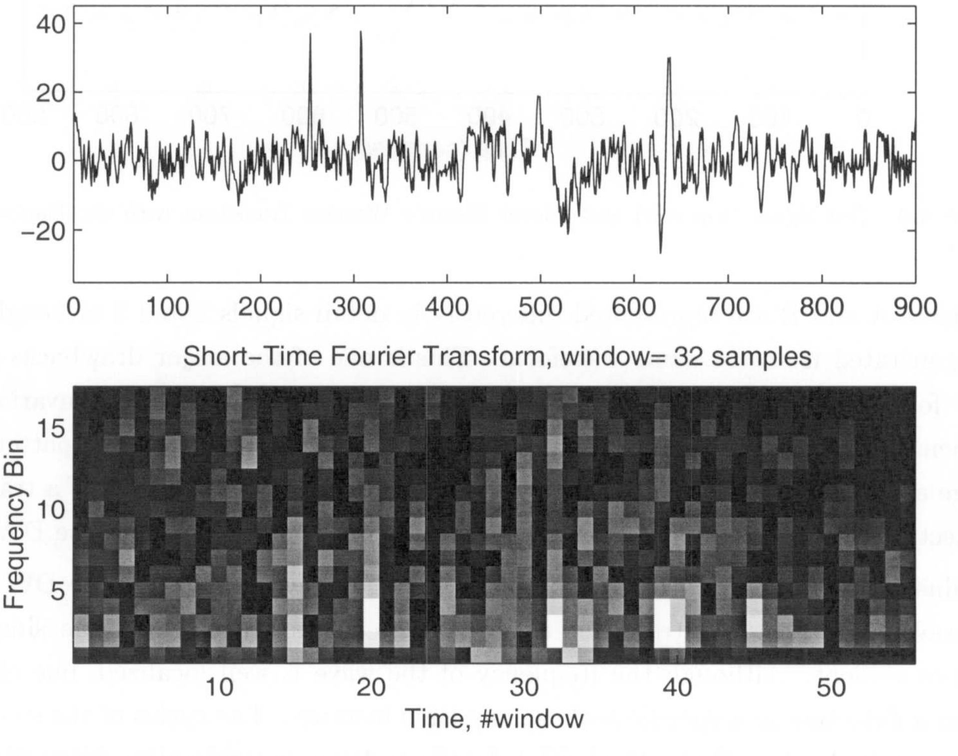

Figure 4.4 Test signal (top row) and Short-Time Fourier Transform with window width 32 samples.

4.2.4 Discrete Wavelet Transform

The DWT (Daubechies-2 wavelet) was applied to the same test signal (Fig. 4.5). Four levels of detail are shown with '1' representing the smallest scale, '4' representing the largest scale, and the dashed curve corresponding to the approximation. All transients are clearly visible in the detail signals. Since smaller scales correspond to shorter windows and larger scales to wider windows, all transients can be detected through a single transform. Transient C is represented by the approximation signal. Significantly,

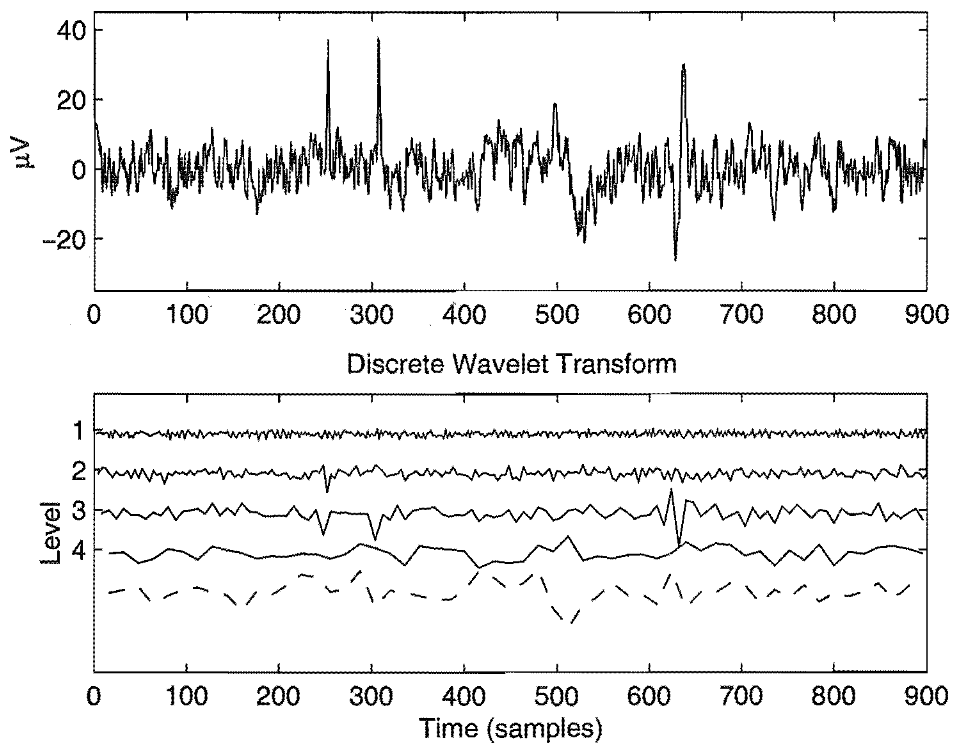

Figure 4.5 Test signal (top row) and 4-level Discrete Wavelet Transform with the Daubechies-2 wavelet.

transients A and B are represented differently in detail signals 2 and 3 although they were generated from the same waveform. This is one of the major drawbacks of the DWT for the EEG transient detection application: it is not translation-invariant. A transient in the signal causes a specific pattern in the transform but this pattern will change as soon as we shift the signal by a couple of samples. Thus, even if a transient is detected, it is not possible to extract a transient-specific feature from the DWT.

Illustration of the lack of translation-invariance is even clearer in the DWT of a sine wave (Fig. 4.6). A strong interference between the wavelet and the sine wave becomes evident. Although the frequency of the wave is well localized, one can not rely on a detection in a specific scale at any time instance. The cycles of the sine wave are represented primarily by the level-4 details and the approximation intermittently, forming what is in effect an interference pattern. Note the high amplitude of the wavelet

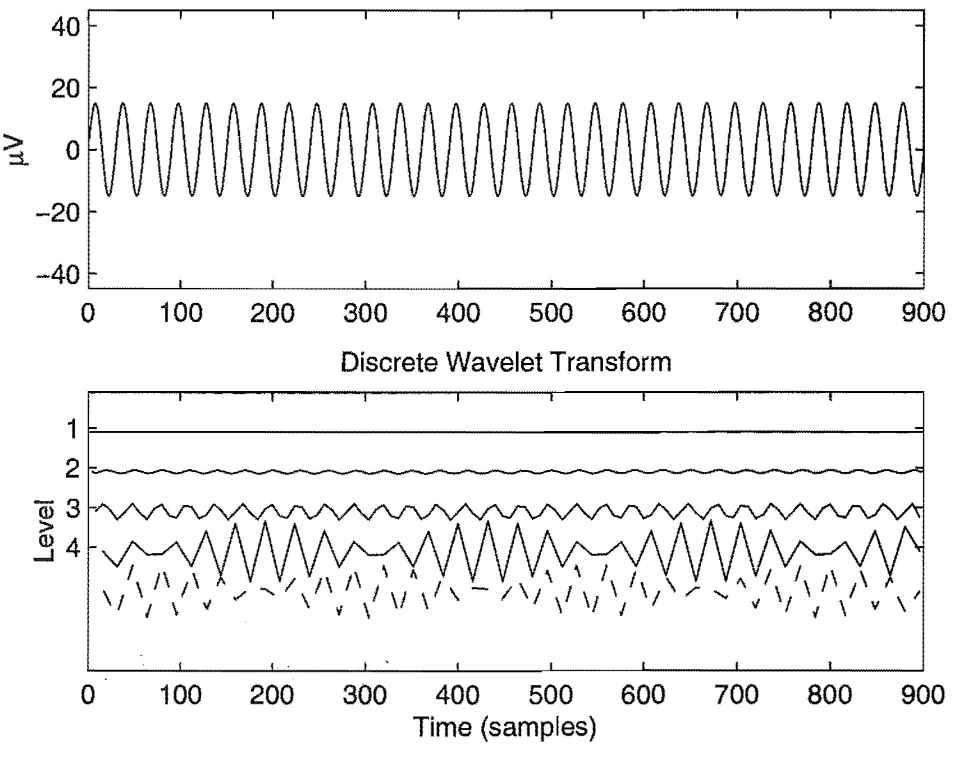

Figure 4.6 Sine wave (top row) and DWT with Daubechies-2 wavelet.

coefficients when they are in phase with the wave. At any one time one could not rely on a detection pattern for a cycle or half-cycle of the wave.

4.3 CONTINUOUS WAVELET TRANSFORM

Unfortunately, it is not possible to design a transform which is translation-invariant, orthogonal and invertible at the same time [Strang and Nguyen 1997]. Of these attributes, however, translation-invariance is essential for a sensitive detection system. It is shown here that the CWT has advantages over the DWT in the detection of a transient waveform.

The CWT of a signal can be considered as the output of a series of band-pass filters. The analyzing wavelet represents a prototype filter (commonly a modulated window function, see previous chapter) which is re-scaled to a series of scales to cover the relevant spectrum. No downsampling occurs as does in the DWT. The DWT provides only one scale per octave (corresponding to each level), whereas the CWT can explore scales and frequencies in smaller steps. Scale is inversely proportional to frequency with a proportionality factor which is dependent on the wavelet filter. The output of a CWT is commonly referred to as a *scalogram* since its vertical axis relates to scale rather than frequency. Some authors define the scalogram as the power (squared magnitude) of wavelet coefficients [Quian and Chen 1996]. For a complex-valued wavelet filter

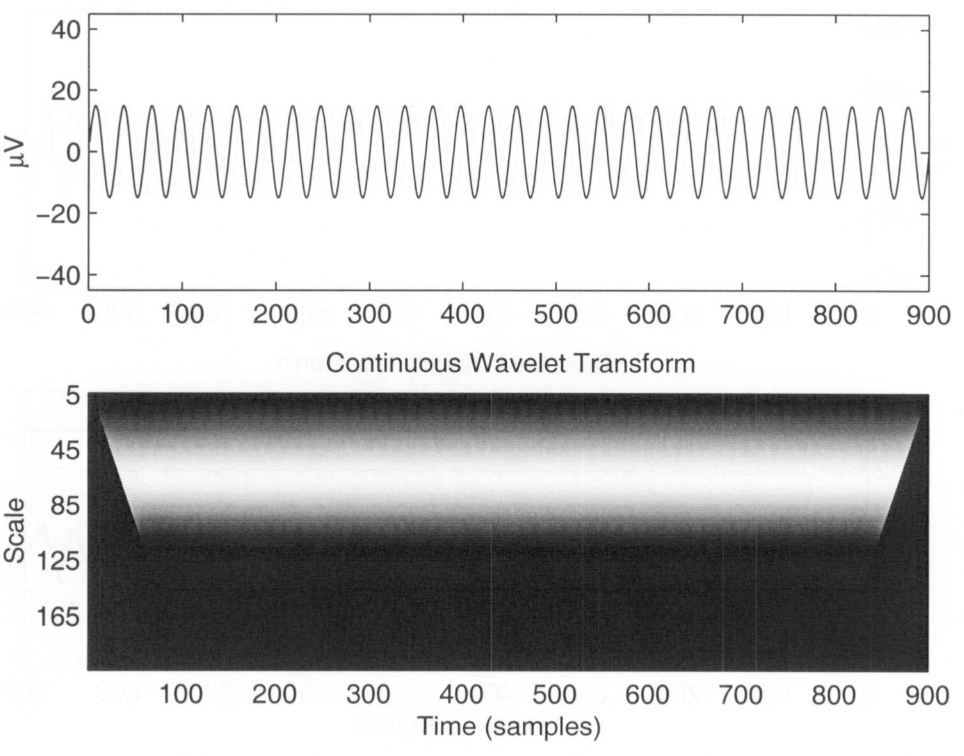

Figure 4.7 Modulus of the Continuous Wavelet Transform of a sine wave using the Senhadji wavelet as analyzing wavelet. Note that the transform is virtually independent of time.

the modulus (magnitude) of the resulting complex wavelet coefficients is commonly depicted in the scalogram.

Fig. 4.7 shows the scalogram of a sine wave. The Senhadji wavelet (modulation $k = 2$) was used as an analyzing wavelet (Eq. 3.9). A discrete set of scales is used, rising linearly from 5 to 201 samples[1]. Interference between the wavelet and the wave is barely noticeable. The scalogram shows a ridge structure that is fairly constant over time because the complex wavelet filter separates magnitude and phase of the oscillation. The complex wavelet filter is equivalent to the *complex demodulation* method that causes a shift of the spectrum, followed by a low-pass filter [Unser 1994].

The DWT (previous section) and the CWT of the sine wave show one of the weaknesses of the DWT. Since the filters of the DWT are real-valued the phase-interference with the oscillation cycles of the sine wave cannot be suppressed. However, this interference can be eliminated by the use of a complex wavelet filter with the CWT.

4.3.1 Normalization and Frequency Bias

In the example presented in Fig. 4.7, the strongest response of the CWT to the sine wave can be seen near scale 65. This is confirmed by the plot of the modulus of the

[1]Scales for the Senhadji wavelet or any binomial kernel wavelet are given in samples.

CWT at sample 450 shown in Fig. 4.8 (solid line). However, since the sine wave was generated by

$$s(k) = 15 \cdot \sin(2\pi k\, 30/900) \tag{4.1}$$

and the analyzing wavelet accommodates two oscillation cycles on its compact support, one would expect to find the peak response at scale 60. As mentioned in the previous

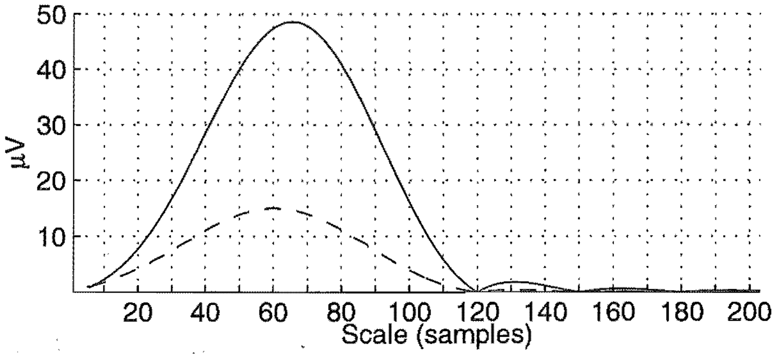

Figure 4.8 Vertical cross section of the CWT in Fig. 4.7 at sample 450. Solid line: energy-normalization, dashed line: true amplitude normalization.

chapter, wavelet filters are commonly normalized to a unit power. This normalization method will be called 'power normalization' in this work. It was originally chosen for the DWT so that the overall signal power is maintained. Unfortunately, this normalization results in a biased estimate for the frequency and the amplitude of a signal component. With a different normalization factor ($1/a$ rather than $1/\sqrt{a}$ in Eq. 3.3 on p. 32) the peak in the CWT corresponds to the right frequency.

In addition, a good estimate for the amplitude of an oscillating component may be required. This can be achieved by normalizing the wavelet filter such that its DFT has a maximum magnitude of 2. In this way, an unbiased estimate for scale *and* amplitude is obtained (Fig. 4.8, dashed line). This normalization is henceforth called 'true amplitude normalization'.

The CWT is now applied to the test signal of the previous section. Both normalization methods are applied: 'power' and 'true amplitude' (Fig. 4.9). All four transients can clearly be seen in the scalograms of the test signal. The response to transients A and B is now very similar, although a wave coincides with transient B leading to a larger peak in the scalogram compared to transient A. Even transient features of the background can be traced in the dendrite structure of the scalogram.

The true amplitude normalization emphasises small scales and high frequencies. Thus the large scale transient C is assigned a lower relative magnitude with respect to the other transients.

The CWT offers a more detailed, redundant, translation-invariant representation

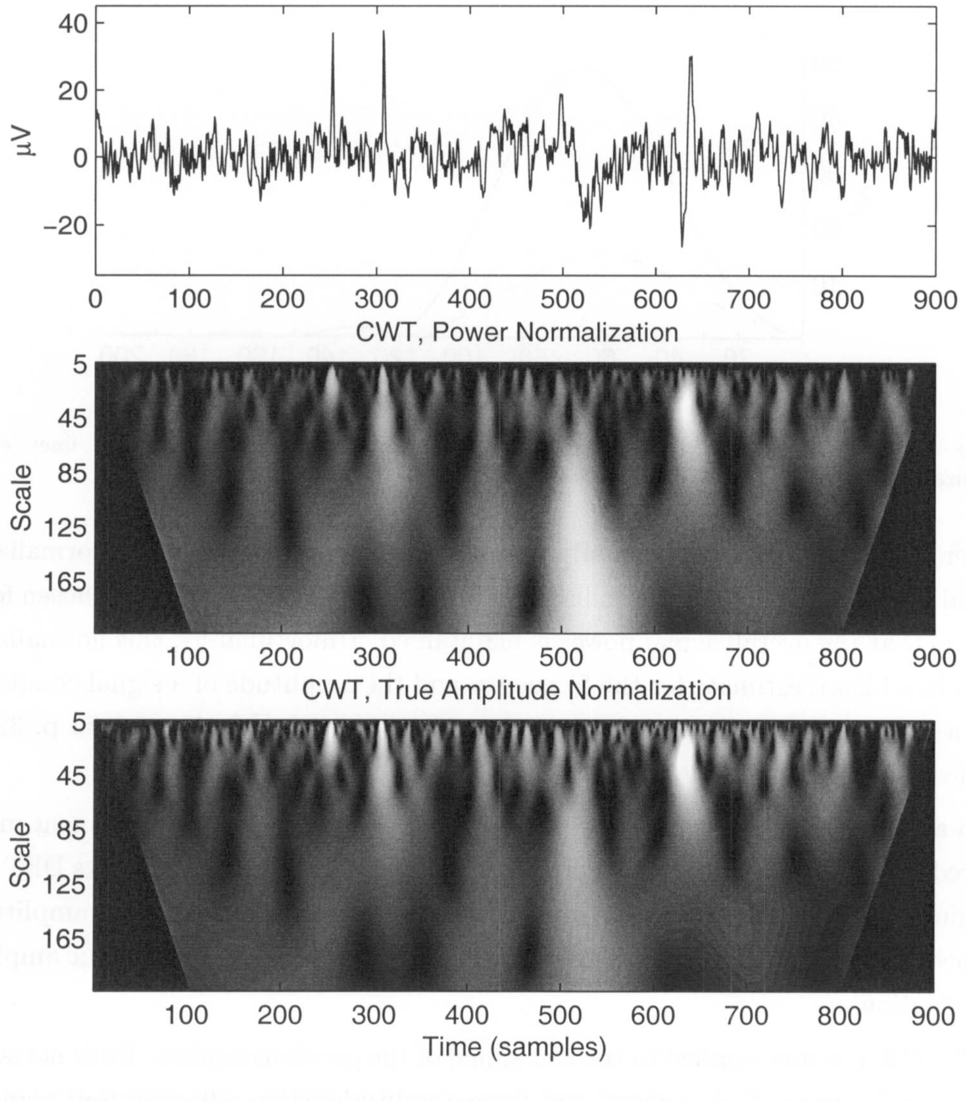

Figure 4.9 Magnitude of the continuous wavelet transforms of the test signal. Analyzing wavelet: Senhadji wavelet with modulation $k = 2$. Two normalization methods were used: (middle) power normalization (factor $1/\sqrt{a}$ as defined in Eq. 3.3) and (bottom) true amplitude normalization (factor changed to $1/a$).

of the data. Unfortunately, its computational complexity is orders of magnitude larger than that of the DWT and there is no simple method to invert the CWT like the DWT. A possible approach to achieve relative translation invariance is to oversample the DWT [Sari-Sarraf and Brzakovic 1997]. For a review of such variants with application to analyzing seizure EEGs see Schiff *et al.* [1994b]. Only the CWT allows for the application of a complex modulated window function as an analyzing wavelet, which results in a continuous estimate of signal power in a frequency band without phase interference.

4.3.2 Multiscale Edge Extraction

The foregoing sections have shown that oscillations are represented by horizontal structures ('ridges') while transient waveforms correspond to vertical structures in the CWT. In order to classify transients, the features of the vertical structures in the CWT can be analyzed in detail. Transients could be peaks, pulses or single step-like features that will be called 'edges' here.

Edges are well localized in time but poorly localized in frequency. Thus they cannot be fully represented by a single point in the scalogram as could the transients in the previous test signal. A strategy to extract the features of a transient is therefore to trace temporal maxima in the scalogram across scales. These traces have been termed *fingerprints* or *multiscale edges* [Witkin 1983, Mallat and Zhong 1992, Mallat and Hwang 1992]. A test signal similar to the one used by Mallat and Zhong [1992] and Mal [1999] was generated (Fig. 4.10 top). The test signal features several sharp transients and edges.

4.3.2.1 Real Wavelet

Commonly the DOG1 wavelet [Witkin 1983] or a similar wavelet is used in this context [Mallat and Zhong 1992, Mallat and Hwang 1992, Evertsz *et al.* 1995, Berkner and Wells 1999]. Here the imaginary part of the psi_1 wavelet was applied (Fig. 4.10 bottom). The psi_1 wavelet was used so that a comparison between the results obtainable with a real-valued and complex-valued wavelet can be made. A harmonic progression of scales with 18 scales per octave was used. As the analyzing wavelet was real, both local temporal maxima (black dots) and local minima (white dots) are marked in the scalogram.

For each major transient either (a) a single multiscale edge (samples 480, 750, 850) or (b) groups of three multiscale edges (near samples 180, 240 and 1050) are found in the scalogram.

At small scales, the two steps in the signal near samples 180 and 240 are represented by two individual multiscale edges. Those multiscale edges merge near scale 400, so at

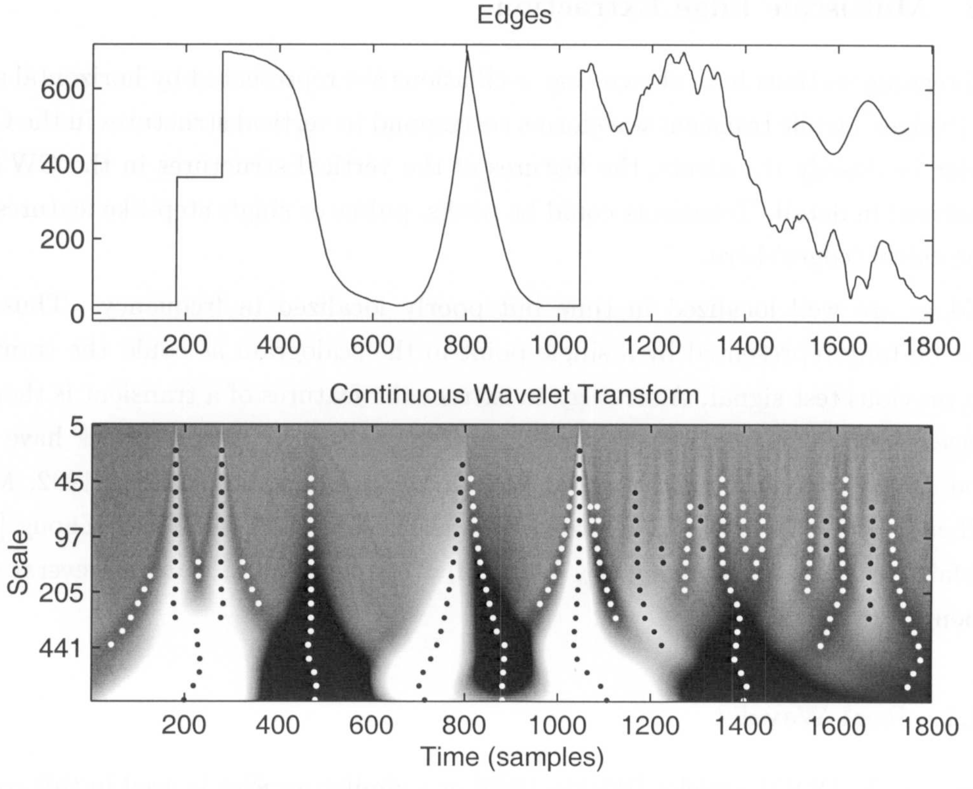

Figure 4.10 A test signal featuring sharp transients (top) and corresponding scalogram (true amplitude scaling, 18 scales per octave). Local temporal maxima and minima are marked with black and white dots, respectively. The analyzing wavelet (imaginary part of psi_1) is shown in top right corner. The test signal was padded with zeros on both sides in calculation of the scalogram.

a coarser scale both steps are 'seen' as a single edge. Neighbouring multiscale edges of the same polarity merge at a certain scale, since the detail information to differentiate between them is no longer available at that scale.

In contrast, the distance of neighbouring multiscale edges of opposite polarity grows with increasing scale. An example of this behaviour can be found near sample 800.

Several features of an edge can be derived from the amplitude decay of the CWT along the corresponding multiscale edge. Mallat and Hwang [1992] estimated the local Lipschitz regularity α of edges with an iterative method. Evertsz *et al.* [1995] also determined the local Hölder exponent f. Alternatively, the scale at which the largest amplitude is found and the scale at which the multiscale edge ends have been proposed as additional features [Evertsz *et al.* 1995, Berkner and Wells 1997, Berkner and Wells 1999] .

4.3.2.2 Complex Wavelet

In Fig. 4.11 the full complex wavelet (psi_1) has been applied to the previous test signal. The scalogram shows the modulus of the CWT coefficients with local temporal maxima marked by dots. Now only one multiscale edge is found for each transient. Here the rising and falling edges of the sharp peak near sample 800 are represented by a single multiscale edge. Thus, with a complex-valued wavelet, a peak in the signal is generally processed as a single entity like an edge.

The CWT estimates the local energy content of a signal. Only with a complex wavelet can this estimate be made independent from the phase of the signal. If for example the phase of an edge has been distorted by some filtering process or is generally unknown, the complex wavelet is able to pick up the full amount of energy associated with the transient. The phase of wavelet coefficients along the multiscale edges is available as additional information about the transient. If noise is present in the signal (as in samples 1100 to 1800) it becomes difficult to trace multiscale edges. Even multiscale edges derived from a complex-valued CWT merge at some scale when moving from small scale to large scales. However, the rules for merging are different. Even multiscale edges of a rising and a following falling edge may merge to cover a peak. This phenomenon can be observed, for example, near samples 220 and 500. In Fig. 4.10 there are two separate multiscale edges of opposite polarity at scale 450 (near samples 220 and 300). However, in Fig. 4.11 the same multiscale edges merge near scale 450 (sample 350), representing a large scale positive peak in the signal. Thus, with a complex wavelet the concept of multiscale edges can be generalized to cover peaks and transients.

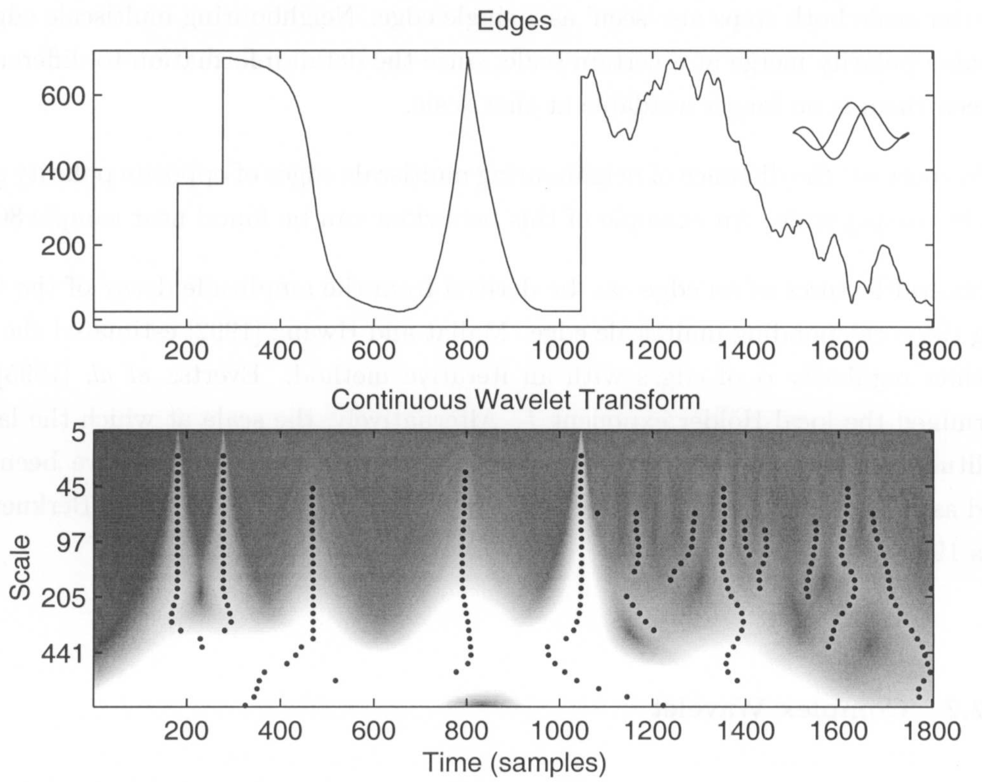

Figure 4.11 Scalogram of sharp transients test signal (as in Fig. 4.10) but analyzed with a complex wavelet (shown in top right corner). Temporal modulus maxima were marked with black dots.

4.3.3 Ridge Tracking

The previous section showed how features of transients can be extracted from the temporal maxima of the CWT. Local maxima with respect to scale may also be present in the CWT. These horizontal ridge structures can be used to estimate the instantaneous frequency of an oscillation. As an example, a linear chirp was embedded in Gaussian white noise (SNR= -5 dB). The chirp signal is shown in Fig. 4.12 together with the corresponding CWT. The Senhadji wavelet with modulation $k = 3$ was applied. Local maxima of the CWT with respect to scale are marked with black dots. For moderate noise conditions, as between samples 100 and 400, the ridge provides a reasonable direct estimate of the instantaneous frequency. The frequency of the chirp increases linearly with time, thus the scale of the chirp follows the function $1/x$. At smaller scales (e.g., near sample 700) the noise interferes with the chirp signal making the ridge discontinuous.

In severe noise situations the ridge can be searched in a high-dimensional space of curves with iterative methods as proposed by Carmona *et al.* [1997]. A larger relative bandwidth of the wavelet would improve the noise suppression and frequency resolution but would also limit the speed of change in frequency traceable with the CWT.

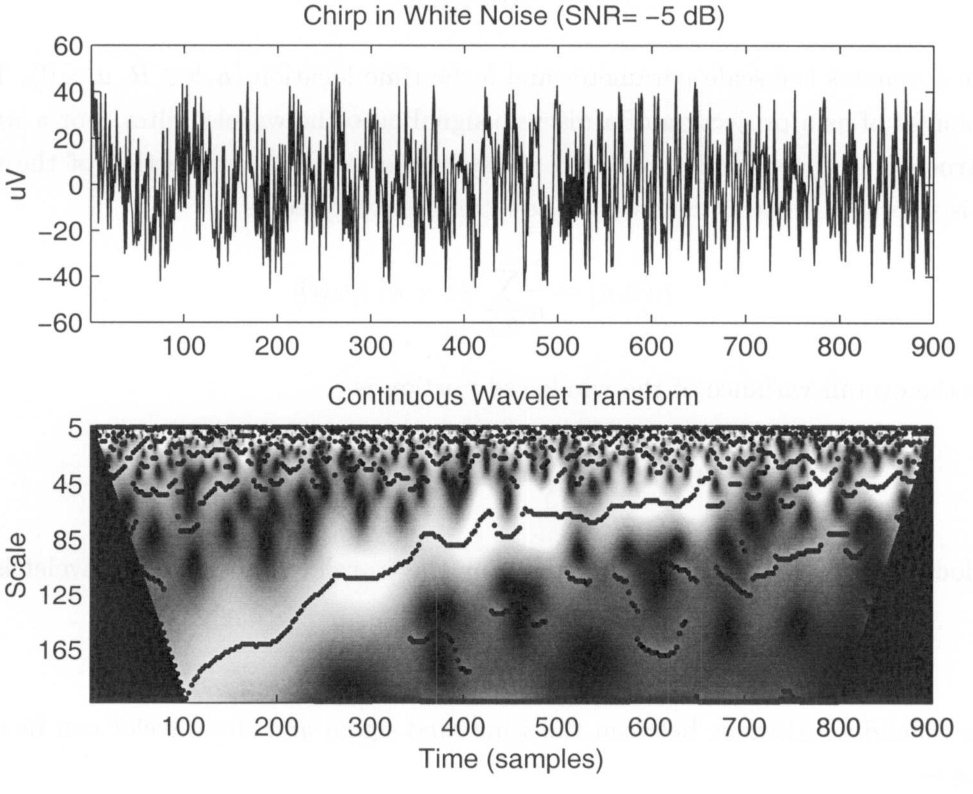

Figure 4.12 Linear chirp in Gaussian white noise (SNR= -5dB) and corresponding CWT. Local scale maxima in the CWT were marked with black dots. The Senhadji wavelet with modulation $k = 3$ was used.

4.3.4 Distance Subtraction Method

The smooth nature of features in the scalogram can be considered a consequence of the lower limit of the time-frequency bandwidth product (Eq. 3.16). This means that no single time-scale point in the scalogram could have a non-zero magnitude without influencing its neighbourhood. A method is proposed here which attempts to improve the resolution of the scalogram.

The CWT for discretely sampled signals (Eq. 3.4) can be re-written as

$$\mathrm{CWT}(k,a) = \frac{1}{\sqrt{a}} \sum_{l \in Z} s(l+k)\, \psi_a^*(l) \tag{4.2}$$

where a denotes the scale parameter and b the time location ($a, b \in R$, $a > 0$). It can be thought of as a projection of a discrete signal onto the wavelet filter. For a wavelet constructed as a complex modulated window, the mean of the portion of the signal that is selected by the translated window can be expressed as

$$\mu(k,a) = \frac{1}{a} \sum_{l \in Z} s(l+k)\, |\psi_a(l)| \tag{4.3}$$

while the overall variance of the windowed portion is

$$\sigma_s^2(k,a) = \frac{1}{a} \sum_{l \in Z} \left| \Big(s(l+k)\, |\psi_a(l)| \Big) - \mu(k,a) \right|^2 . \tag{4.4}$$

The local variance of the signal expressed by the translated and scaled wavelet is

$$\sigma_w^2(k,a) = \frac{2}{a} \cdot |\mathrm{CWT}(k,a)|^2 \tag{4.5}$$

so the Euclidian distance between the windowed signal and the wavelet can be determined as

$$d(k,a) = \sqrt{\sigma_s^2(k,a) - \sigma_w^2(k,a)} . \tag{4.6}$$

The distance is small if the windowed signal matches the frequency of the scaled wavelet and grows larger if other frequencies are predominant the windowed segment. By subtracting this distance from the CWT

$$\mathrm{CWT}_d(k,a) = |\mathrm{CWT}(k,a)| - d(k,a) . \tag{4.7}$$

a *distance subtracted CWT* can be defined. An example with the original CWT and the distance corrected CWT of a chirp signal is shown in Fig. 4.13. The important property of this representation is a stronger focus on actual components with a more rapid decay around them. The amplitude is not preserved by this method since the term

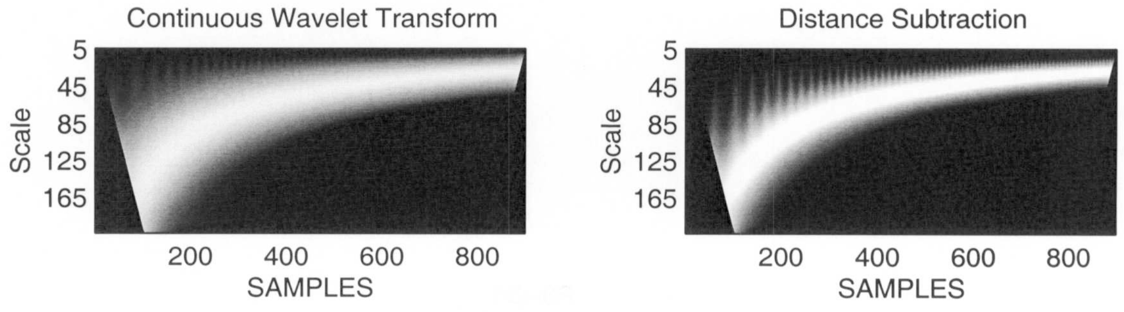

Figure 4.13 CWT of linear chirp (left) and CWT after distance subtraction (right). The Senhadji wavelet with modulation $k = 2$ was used.

$d(k, a)$ is subtracted but a stronger definition of structures with respect to background is achieved. Unfortunately, ripples appear on one side of the ridge in the scalogram due to phase interference. The ripples are in phase with the zero-crossings of the chirp. They correspond to the local high frequency content of the maximal slopes of the chirp signal.

4.4 MATCHING PURSUIT

In the Fourier transform, frequencies are extracted by changing the modulation of the basis function. In the CWT the scale of the basis function is changed instead. A wavelet with a fixed modulation is applied which will be sufficient for many applications. The variety of features in the EEG is, however, large and includes oscillatory elements just like transients. A good example are alpha waves [Nunez 1981]. For example, Fig. 4.14 shows a bipolar recording from the occipital region which contains several alpha spindles. In the scalogram the waves are clearly visible but their frequency cannot be determined very well. It could be somewhere between 7.5 and 12.5 Hz. Isolated transients, like the one near sample 850, are also picked up by the $10\,\mu$V contour line in the scalogram.

In Fig. 4.15 the modulation of the analyzing wavelet has been changed from $k = 2$ to $k = 8$. Now one can obtain the alpha frequency with reasonable precision as being close to 9 Hz. At the same time the transient event near sample 850 disappears from the scalogram. It would clearly be desirable to combine both analyzing wavelets in a single analysis. Mallat and Zhang's [1993] matching pursuit, introduced in section 3.6, is not limited to a single analyzing wavelet. It has, instead, a dictionary of waveforms, all of which are normalized to unit energy. The signal is filtered with all waveforms and the sample with the largest magnitude among all responses is chosen. An instance of the corresponding waveform is translated and scaled appropriately and is called a time-frequency atom of the signal. The atom is subtracted from the signal and added to a reconstruction signal. The procedure is repeated until the difference between original and reconstruction becomes sufficiently small.

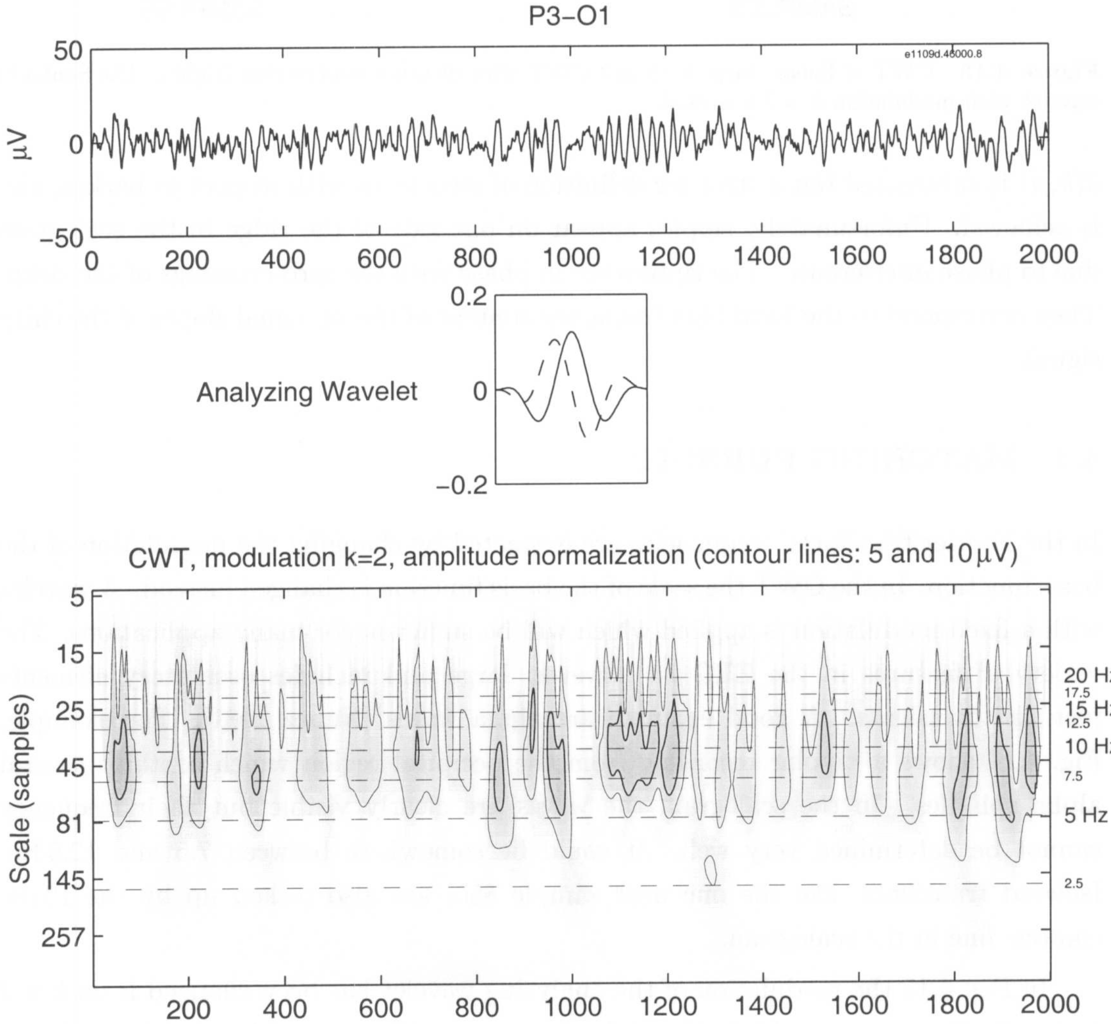

Figure 4.14 Series of alpha waves in bipolar EEG recording (10 s, 200 Hz sampling rate) and corresponding scalogram. Senhadji wavelet, modulation $k = 2$, 6 scales per octave. A frequency scale is added on the right hand side of the scalogram. Sharp edges of the alpha waves reach into higher frequencies.

In Fig. 4.16 the same EEG trace as in the previous figures was analyzed with the matching pursuit method. The dictionary covered 71 waveforms including 20 scales (4 per octave) and 4 modulations ($k = 1, 2, 4, 6$) of the Senhadji wavelet. The reconstruction signal is plotted after 5, 10, 15, and 20 iterations. After 5 iterations an alpha

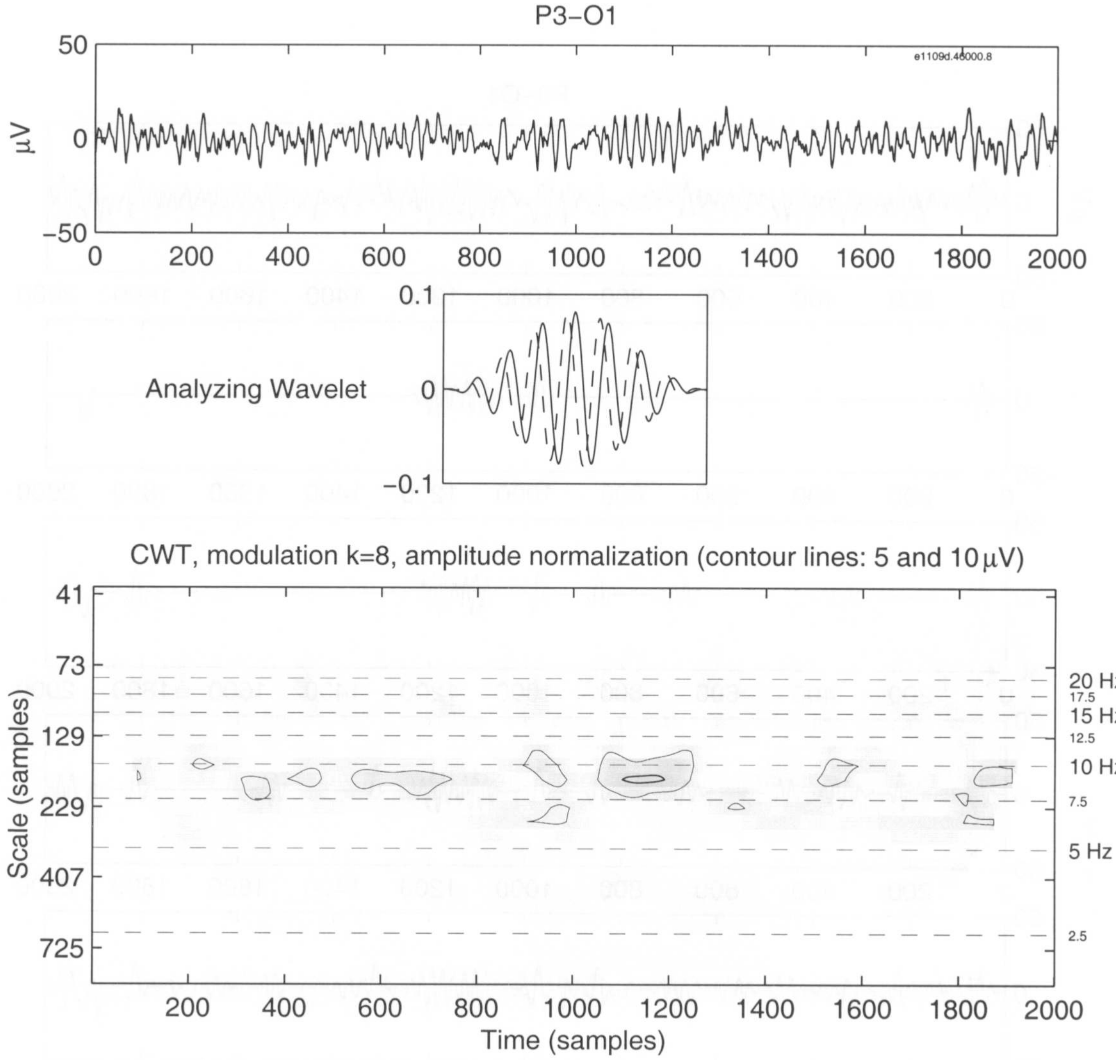

Figure 4.15 Series of alpha spindels in bipolar EEG recording and corresponding scalogram (modulation $k = 8$).

wave is found near sample 1150 in the reconstruction. It is represented by two adjacent time-frequency atoms. Therefore, the most central peak of the wave is attenuated in the reconstruction. The wave would be better represented by a larger modulation (e.g., $k = 8$ or 10). The isolated transient near sample 850 appears in the reconstruction after 10 iterations but its characteristic sharp edges are lost. The sharp edge details would eventually be picked up by atoms in further iterations. As one could see in the previous section, edge details are distributed across scales. With matching pursuit, a sharp edge would be decomposed into several atoms.

Matching pursuit is the best way to see both 'the forest and the trees' in a signal but it is a computationally very expensive algorithm. In this example, the signal had to be filtered 1420 times to extract 20 time-frequency atoms and, hence, is an order of magnitude slower than the CWT. For the application of matching pursuit to sleep waves see Durka and Blinowska [1996] or Zygierewicz *et al.* [1999].

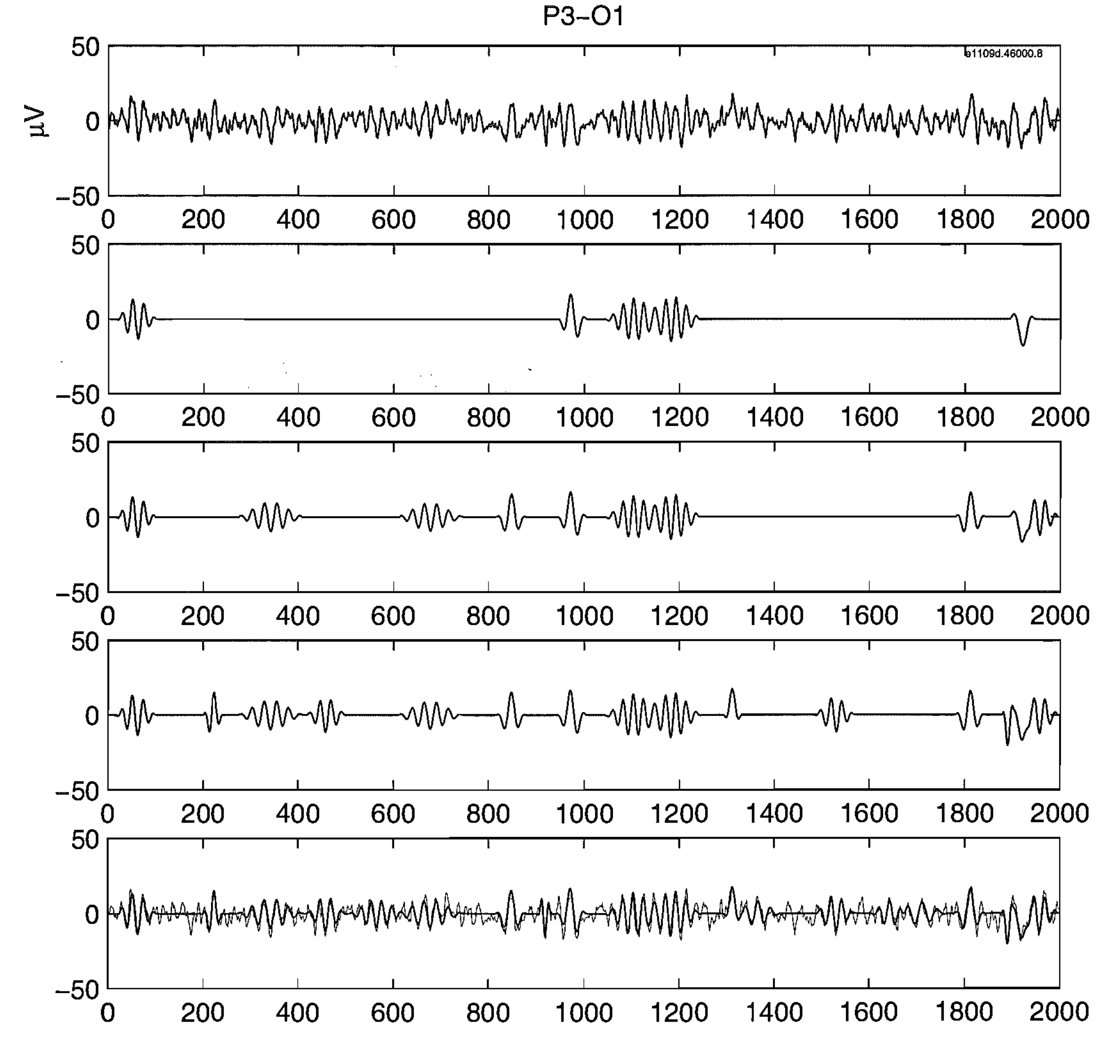

Figure 4.16 An example of matching pursuit: Original signal (top row) followed by reconstruction signal after 5, 10, 15, and 20 iterations. The original signal is superimposed as thin line in the bottom row.

Alternatively, the waveform dictionary can be derived from actual data, or be combined with actual data such as ETs or artifacts. Thus, given a sufficiently large amount of computational power, a signal could be decomposed into atoms representing ET and/or artifacts.

4.5 COMPARISON TO TIME-FREQUENCY ANALYSIS

The short-time Fourier transform, DWT and CWT can be regarded as *linear joint time-frequency distributions.* As counterparts to these, there is a range of *bilinear joint time-frequency distributions* such as the *Wigner-Ville distribution* or the *Choi-Williams distribution.*

They are described in the following subsections and compared to the CWT.

The WVD was originally developed for the area of quantum mechanics by Wigner in 1932 and later introduced to signal analysis by Ville. Most bilinear time-frequency distributions are based on the WVD [Quian and Chen 1996].

4.5.1 Wigner-Ville Distribution

The WVD of a signal $s(t)$ can be defined as

$$\mathrm{WVD}(t,f) = \int_{-\infty}^{\infty} s(t+\tau/2) \cdot s^*(t-\tau/2)\; e^{-i2\pi f\tau}\, d\tau \tag{4.8}$$

where s^* denotes the complex conjugate of s. Since two time instances of the signal are multiplied in the convolution, the VWD is a *bilinear* distribution as opposed to the Fourier Transform, CWT or DWT. There are several attractive properties of this distribution such as *translation and modulation covariance*

$$s'(t) = s(t+t_0) \rightarrow \mathrm{WVD}_{s'}(t,f) = \mathrm{WVD}_s(t-t_0,f) \tag{4.9}$$

$$s'(t) = s(t)e^{i2\pi f_0 t} \rightarrow \mathrm{WVD}_{s'}(t,f) = \mathrm{WVD}_s(t,f-f_0) \tag{4.10}$$

time and frequency marginal properties

$$\int_{-\infty}^{\infty} \mathrm{WVD}(t,f)df = s(t) \tag{4.11}$$

$$\int_{-\infty}^{\infty} \mathrm{WVD}(t,f)dt = S(f) \tag{4.12}$$

with $S(f)$ being the FT of $s(t)$ and *energy conservation*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{WVD}(t,f)dtdf = \int_{-\infty}^{\infty} s(t)^2 dt\,. \tag{4.13}$$

The WVD of a linear chirp in white Gaussian white noise (SNR= -5 dB) shows a narrow ridge representing the chirp which is commonly referred to as *auto-term* (Fig. 4.17). Auto-terms relate to actual components in the signal. The auto-term is surrounded by a large amount of high-amplitude noise. Although the frequency distribution of the noise is broad, the amplitude of the noise in the WVD is as large as that of the chirp. The major disadvantage of the WVD is that it features interference patterns between

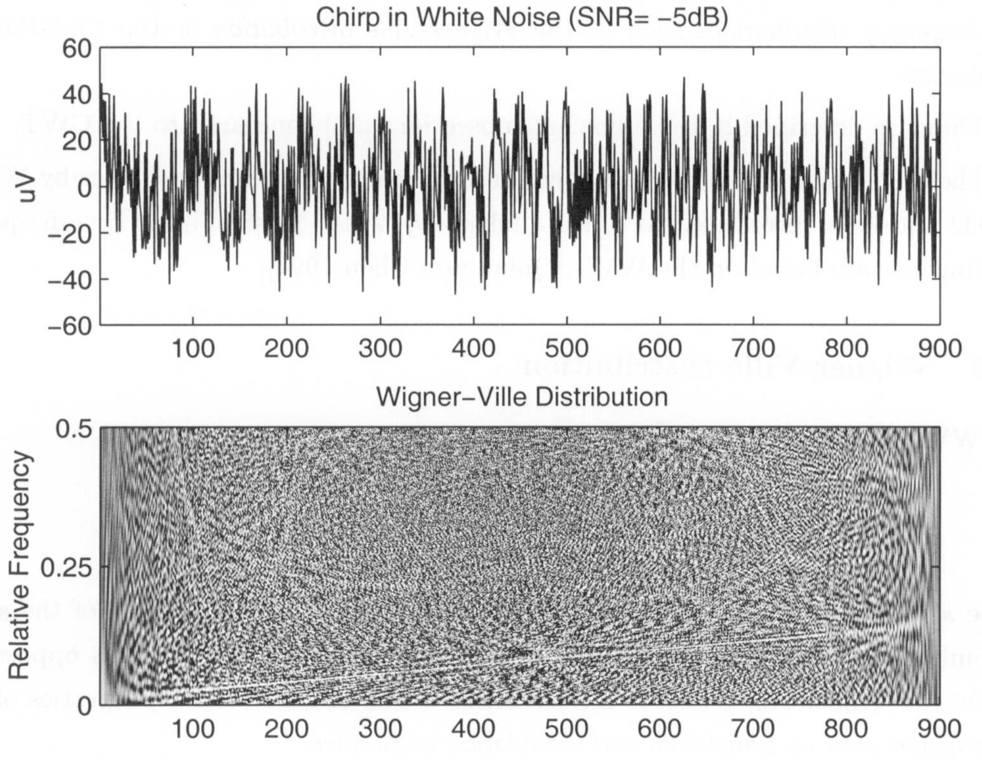

Figure 4.17 Wigner-Ville distribution of linear chirp in white Gaussian noise (-5 dB).

auto-terms called *cross-terms* of the same magnitude as the auto-terms. If the time-frequency distribution of a signal has a complicated structure it may be impossible to differentiate auto-terms from cross-terms.

To study the transient representation properties of the WVD, the transient test signal was analyzed (Fig. 4.18). All transients (A, B, C and D) can be identified in the WVD. Transients A and B (near samples 250 and 310 respectively), lead to narrow auto-terms that cover a wide frequency range. They corresponds to the wide spectra of these transients as found in their FT (Fig. 4.2a, p. 46). The same correspondence is evident for transients C and D (near samples 520 and 630, respectively). In addition, a typical cross-term is particularly prominent midway between these auto-terms (near sample 580). It should be noted that adding the temporally isolated transients caused an interference in the FT similar to the cross-terms found in the WVD (Fig. 4.2b, p. 46).

A detector based on the WVD would have to integrate the energy of transients A and B, which is wide-spread in the frequency domain. In comparison, the respective response of the CWT (Fig. 4.8b, p. 51) is much more focused. However, the CWT can only be focused on a certain kind of transient, if a wavelet with a matched relative

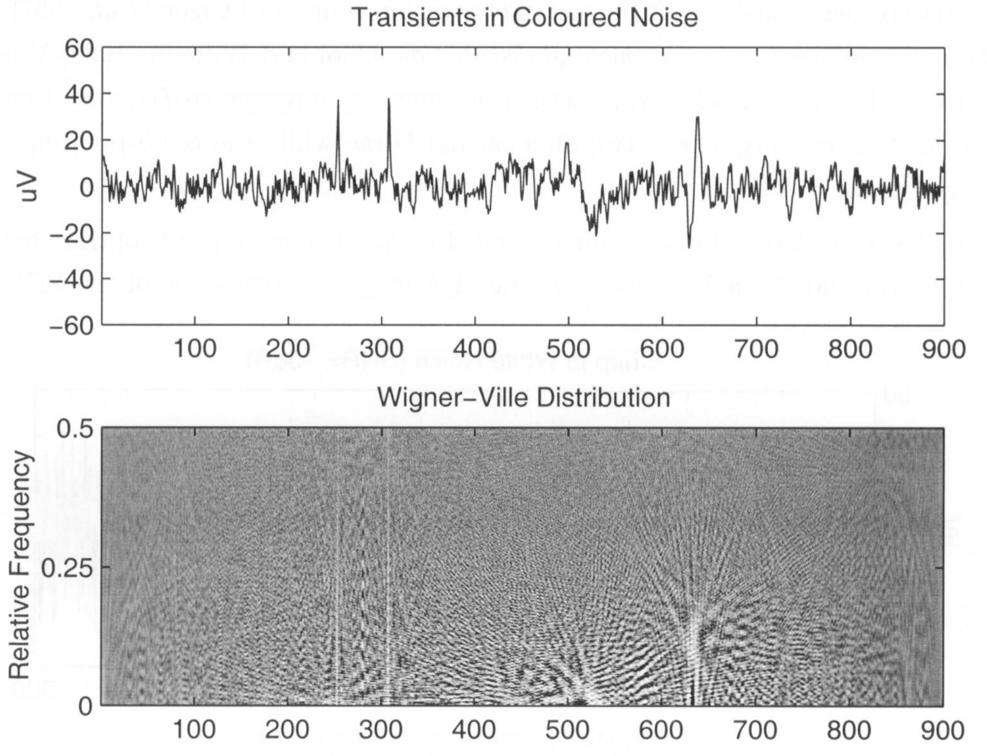

Figure 4.18 Wigner-Ville distribution of transient test signal.

bandwidth Q is used.

4.5.2 Choi-Williams Distribution

A variety of time-frequency distributions have been proposed to eliminate the cross-terms while maintaining the attractive properties of the WVD. Cross-terms oscillate between positive and negative values, while auto-terms are positive only. Therefore, many approaches have been made to smooth the WVD with an appropriate 2D filter or *kernel function*. All of these approaches can be considered to be members of *Cohen's class* of time-frequency distributions

$$\mathrm{C}(t,\nu) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Pi(\tau - t, \varepsilon - \nu)\, \mathrm{WVD}(\tau,\varepsilon)\, d\tau\, d\varepsilon \tag{4.14}$$

with kernel function $\Pi(t, f)$. For the CWD, a Gaussian *parameterization function*

$$f(t,\nu) = e^{-(\pi\nu t)^2/(2\sigma^2)} \tag{4.15}$$

is used. The kernel function

$$\Pi(t,\nu) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\tau,\varepsilon)\, e^{-j2\pi(\tau t+\varepsilon\nu)}\, d\tau\, d\varepsilon \tag{4.16}$$

is the 2D Fourier transform of the parameterization function [Auger *et al.* 1997]. The CWD is also referred to as *exponential distribution* [Choi and Williams 1989, Williams *et al.* 1995]. For $\sigma \to \infty$ the parameterization function converges to $f(t, \nu) = 1$ and the kernel function converges to a two-dimensional Dirac while the corresponding CWD converges to the WVD.

The CWD with $\sigma = 1$ of the chirp signal (Fig. 4.19) shows a good contrast between the auto-term and the noise. However, the high frequency definition of the WVD can

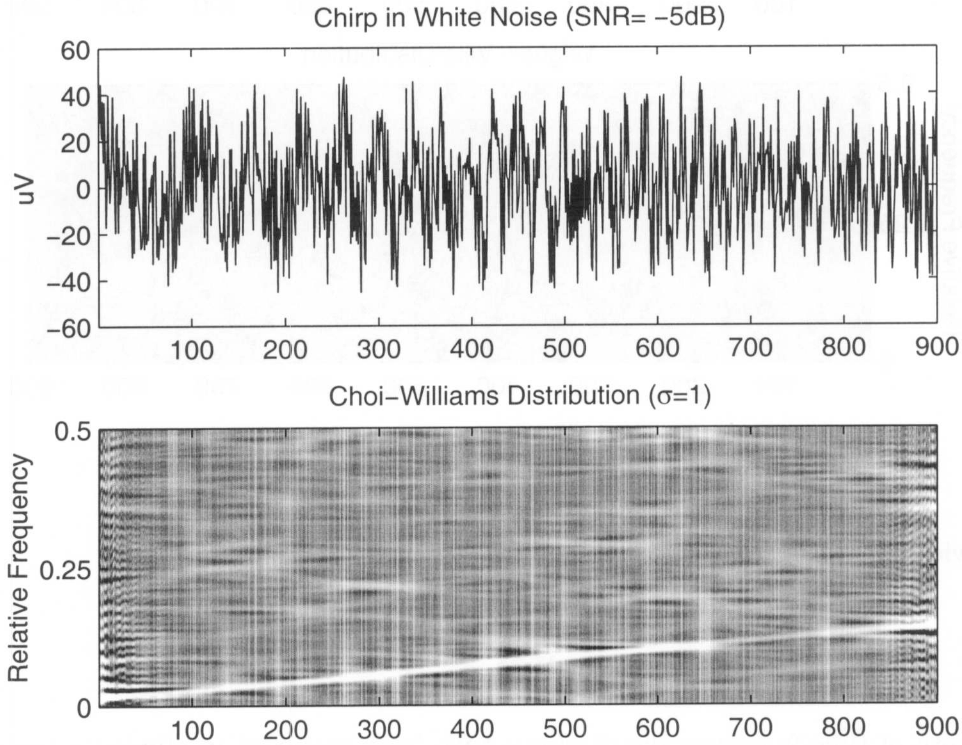

Figure 4.19 Choi-Williams distribution (kernel variance $\sigma^2 = 1$) of linear chirp in white Gaussian noise (-5 dB).

not be achieved. The CWD ($\sigma = 1$) of the transient test signal (Fig. 4.20) highlights the auto-terms while most of the cross-terms (e.g., between transient C and D near sample 580) are suppressed. Transients A and B are well localized in time and clearly defined at all frequencies. For biphasic transient D, two local maxima in the CWD correspond to the positive and the negative peak of the transient. Obviously, a wider low-pass (larger σ in eq. 4.15) would lead to a single local maximum here. A similar situation is found for biphasic transient C. While the negative peak (near sample 520) produces a local maximum, a coinciding clutter of noise changes the appearance of the positive peak suggesting the presence of two distinct frequencies (near sample 500).

The bilinear time-frequency distributions give a less biased analysis of a signal than the CWT. Optimal time-frequency resolution can be achieved if noise is absent. However, for transient representation a short integration window for high frequencies

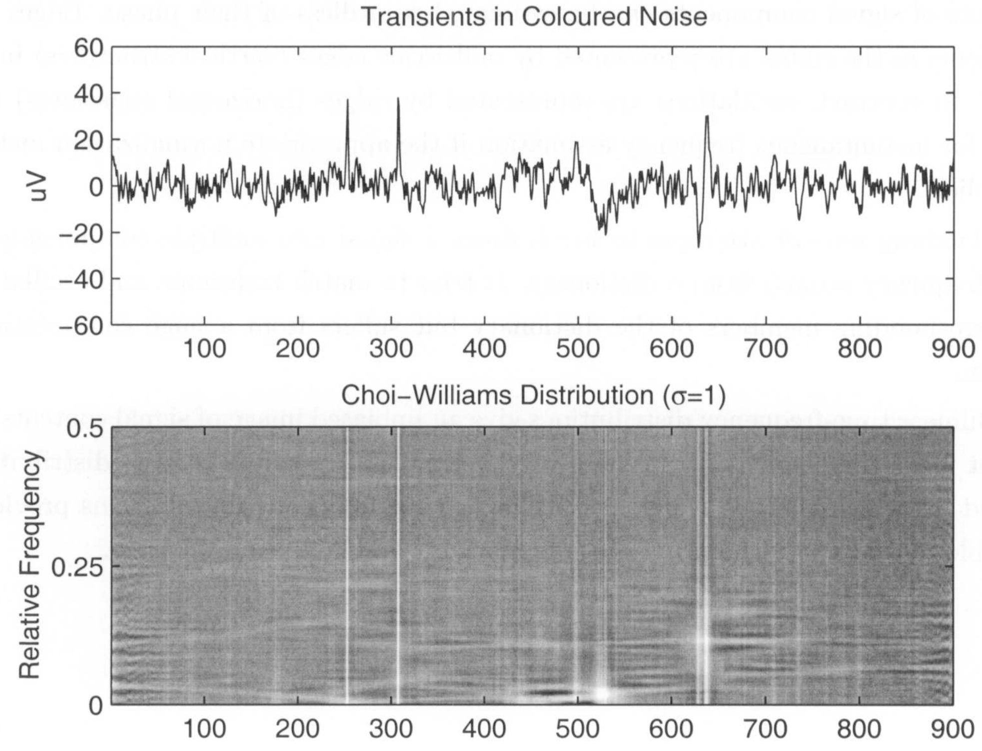

Figure 4.20 Choi-Williams distribution of transient test signal (kernel variance $\sigma^2 = 1$).

and a large integration window for low frequencies (a constant relative bandwidth Q) is desirable. This is the only way to focus the time-frequency distribution such that a *single* local maximum in the distribution is acquired for each transient. The CWT can operate like a matched filter if an appropriate wavelet can be obtained. However, this makes the CWT more biased than bilinear time-frequency distributions.

The modulus of the CWT can be expressed in terms of the WVD

$$|\mathrm{CWT}(a,b)|^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{WVD}(t,f)\,\mathrm{WVD}_{\psi}(\frac{t-b}{a}, af)dt\,df \tag{4.17}$$

where $\mathrm{WVD}_{\psi}(t,f)$ is the WVD of the analyzing wavelet. Thus, the CWT does not belong to Cohen's class but is member of a more general class of time-frequency distributions known as the *affine class*.

4.6 SUMMARY

While the DWT perfectly represents a signal in the sense that it gives perfect reconstruction, it does not lead to a consistent representation of a transient suitable for feature extraction in the transformed domain. This is mainly due to its lack of the translation invariance property. The CWT has this property and in addition allows for the application of complex wavelets. Thus, using the CWT the time-frequency contents of signal components can be estimated regardless of their phase. Edges and transients in the signal are represented by multiscale edges (vertical structures) in the CWT. In contrast, oscillations are represented by ridges (horizontal structures) that allow for instantaneous frequency estimation if the appropriate normalization method is applied.

Matching pursuit attempts to break down a signal into multiple components (or time-frequency atoms) from a dictionary. It tries to match transients and oscillations to corresponding members of the dictionary but suffers from a huge computational burden.

Bilinear time-frequency distributions give an unbiased image of signal contents but do not focus the energy of transients very well unless the affine class of distributions is used, to which the CWT belongs. Bilinear time-frequency distributions provide a valuable tool to study the properties of transients and analyzing wavelets.

Chapter 5

STATISTICAL PATTERN RECOGNITION

5.1 INTRODUCTION

Most decisions are based on previous experience, with some of the most crucial of these being made in medical diagnosis. Medical practitioners can only diagnose a particular condition following extensive training which includes examining many cases for which the presence of the condition is known. Their decision is based on a multitude of features which are more or less indicative for the condition in question. If the features derived from various examinations do not allow for a definitive diagnosis, a probabilistic term such as "suspected", "possible" or "probable" will be part of the doctor's diagnosic report.

A *pattern recognition system* makes the decision to classify a given set of data (a "case") according to an associated pattern that consists of a number of features. The pattern can be considered to be a point in a multidimensional feature space or a *feature vector*. The classification rules define boundaries that divide the feature space into several disjoint regions, the classes.

In statistical pattern recognition, a decision rule is derived from statistical analysis of a training data set. The training data set is meant to represent all available expert knowledge for the decision, so all patterns in the training data set must be labelled according to their class membership. Raudys and Jain [1991] listed the development stages of pattern recognition systems as follows:

- data collection,
- formation of the pattern classes,
- feature selection,
- specification of the classification algorithm and
- estimation of the classification error.

Since the training data set must necessarily be finite, the distribution of features for each class or the *class conditional density functions* can only be estimated with a limited degree of accuracy.

In the work presented here a statistical classifier is used to evaluate transients detected in clinical EEG recordings. Features of each transient are derived from wavelet analysis or analysis of its spatial distribution.

In the following section three statistical classifiers are introduced. To compare the classifiers several measures of performance for the classifiers are presented in section 5.3. A series of simulations follow that illustrate the way the classifiers work. Further simulations with separate training and test sets are used to test the robustness of the classifiers (section 5.4). In section 5.5 a Bayesian approach is proposed for the interpretation of the classifiers' output such that a class membership is expressed by a probability.

5.2 STATISTICAL CLASSIFIERS

Statistical classifiers include such relatively new inventions as *fuzzy logic* [Zadeh 1965, Zadeh 1987, Zadeh 1992] and neural network classifiers like the *self-organizing feature map* [Kohonen 1989, Kohonen 1995]. However, although they date back much further [Fisher 1936, Fisher 1938] they recently received increasing amount of attention as a classical approach to pattern recognition in the field of *artificial intelligence*. They provide a means of implementing *supervised learning*, that is learning to classify patterns from a labelled training set. In this context, the labelled training set represents the expert knowledge built into the classifier.

In the following, a set of cases is assumed to need to be classified. Each case is characterized by a p-dimensional feature vector $X = [x_1, \cdots, x_\mathrm{p}]^T$. Consider two classes of features ω_1 and ω_2 with sample means $\overline{X}^{(1)}$ and $\overline{X}^{(2)}$ and prior probabilities q_1 and q_2, respectively.

Most classifiers presented in the following subsections assume Gaussian distributions for all features but they have been shown to perform well if distributions are non-Gaussian [Raudys and Jain 1991, Vapnik 1995].

5.2.1 Euclidean Distance Classifier

The *Euclidean distance classifier*'s discriminant function

$$g^E(X) = [X - \frac{1}{2}(\overline{X}^{(1)} + \overline{X}^{(2)})]^T \; (\overline{X}^{(1)} - \overline{X}^{(2)}) \tag{5.1}$$

projects the offset corrected input feature vector X onto the direction connecting the means of the two classes. The classification of X is based on the sign of the discriminant function $g^E(X)$:

$$g^E(X) > 0 \rightarrow X \in \omega_1 \tag{5.2}$$

$$g^E(X) \leq 0 \rightarrow X \in \omega_2 \tag{5.3}$$

The Euclidean distance classifier can be applied if the pattern classes are well separated or a simple classification rule is needed [Raudys and Jain 1991].

5.2.2 Fisher's Linear Discriminant

Fisher's linear discriminant assumes a common covariance matrix S for both classes. The discriminant function is given by

$$g^F(X) = [X - \frac{1}{2}(\overline{X}^{(1)} + \overline{X}^{(2)})]^T S^{-1} (\overline{X}^{(1)} - \overline{X}^{(2)}) + \ln(q_1/q_2) \,. \tag{5.4}$$

The term $\ln(q_1/q_2)$ is called the *bias* and it accounts for unequal prior probabilities. As for the Euclidean distance classifier, each input feature vector X is projected onto a fixed direction. However, for Fisher's linear discriminant this direction is parallel to $S^{-1}(\overline{X}^{(1)} - \overline{X}^{(2)})$ which can be shown to maximize the ratio of between-class to within-class variances [Webb 1999]. Fisher's linear discriminant is the most commonly used discriminant function since it has been found to be robust to non-normality of patterns [Raudys and Jain 1991]. However, linearly separable classes[1] can be constructed such that Fisher's linear discriminant results in an error probability close to one [Devroye *et al.* 1996].

5.2.3 Quadratic Discriminant

The *quadratic discriminant function* is based on separate covariance matrices S_1 and S_2 for classes ω_1 and ω_2, respectively

$$g^Q(X) = (X - \overline{X}^{(2)})^T S_2^{-1} (X - \overline{X}^{(2)}) - (X - \overline{X}^{(1)})^T S_1^{-1} (X - \overline{X}^{(1)}) + \ln \frac{|S_2| q_1}{|S_1| q_2} \,. \tag{5.5}$$

For the quadratic discriminant function no global direction of discrimination can be given. The input feature vector X is evaluated in terms of its squared Mahalanobis distances to the respective class means. The *Mahalanobis distance* r between two feature

[1]Two classes are linearly separable if the feature space can be subdivided by a line or hyperplane such that one class falls completely in one subdivision and the other class falls completely in the other subdivision.

vectors X and Y of a set of feature vectors with covariance matrix S is given by

$$r^2 = (X - Y)^T S^{-1}(X - Y) \ . \tag{5.6}$$

Points of equal Mahalanobis distance to some point Y lie on an elliptic curve, whereas points of equal Euclidean distance lie on a circle [Devroye *et al.* 1996]. Thus, the Mahalanobis distance can be understood to emphasize individual features with small variances and de-emphasize features with strong variances.

5.2.4 Multinominal Classifier

A *multinomial classifier* can be used for recognition even if variables take values from a discrete set only. No assumptions are needed for the class conditional distributions. Assume that the jth variable of a feature vector takes m_j distinct values. Thus, there are $m = m_1 m_2 \ldots m_p$ possible patterns or states. Based on the probability p_{ik} of a state $k = 1, 2, \ldots, m$ occuring in class $i = 1$ or 2, the discriminant function is

$$g^H(X) = q_1 p_{1(X^s)} - q_2 p_{2(X^s)} \tag{5.7}$$

where X^s is the index of the state of X. One problem is that the number of states m rises quickly with the dimensionality of the feature vectors even if each variable takes only two distinct values. Therefore it may be difficult to estimate the p_{ik} in practice.

5.2.5 Histogram Classifier

The multinomial classifier can be applied to continuous-valued variables after discretizing them into a set of m_j bins, leading to the *histogram classifier.* The performance of the histogram classifier is dependent on the number of bins and the thresholds separating the bins. Unfortunately, there is no general procedure to obtain the optimal bin thresholds for a given sample [Raudys and Jain 1991].

5.2.6 Bayesian Approach

For each bin of the histogram classifier the Bayes probability for a certain class membership of a feature vector X can be estimated by

$$p(\omega_i|X) = \frac{q_i p_{iX^b}}{\sum_j q_j p_{jX^b}} \tag{5.8}$$

where X^b is the index of the bin corresponding to X. This can be used to obtain a measure of confidence for a decision.

In this work the continuous-valued output of the Euclidian, Fisher or quadratic discriminant functions are discretized to obtain a histogram. Subsequently, the Bayes probability for each histogram bin is obtained to determine the confidence of class membership for one or the other class. This approach is illustrated with the results of simulation B (see below) in section 5.5.

5.3 SIMULATION

The behaviour of the Euclidian, Fisher and quadratic classifiers was studied in two dimensions in simulations A to C, presented below. Their performance and robustness was also be evaluated in a five dimensional feature space in simulations D to F.

Simulations A to C should provide some insight into workings of the classifiers. The same set of data is used to train the classifiers and to evaluate their performance. However, in simulations D to F separate training and test sets are used.

Two Gaussian distributed samples ω_1 and ω_2 of 2D feature vectors were generated. Samples ω_1 and ω_2 were centered around (-1, -1), and (0.5, 0.5), respectively. The three classifiers mentioned in the previous paragraph were applied to three variants of the samples to study the effects of

- equal sample sizes and equal variances of features within each class (simulation A),
- unequal sample sizes and equal variances of features within each class (simulation B), and
- equal sample sizes and unequal variances of features within each class (simulation C).

For equal sample sizes the bias of the Fisher classifier is zero. However, in practice sample sizes may not be equal. By changing the sample sizes but retaining the distribution properties the effect of the bias can be studied.

If variances of features within one class are unequal, that distribution is elliptical. By changing the sample of one class from circular to elliptical, the resulting behaviour of the decision boundaries of the classifiers can be studied.

5.3.1 Performance Indicators

For the study, sample ω_1 was considered to be a widely distributed class of non-targets, while ω_2 was considered to comprise targets to be detected. Performance was measured in terms of the numbers of false detections (FD), rejected non-targets (RN), detected targets (DT) and missed targets (MT). These measures can be sumarized in a *confusion matrix*:

$$\begin{bmatrix} \mathrm{RN} & \mathrm{FD} \\ \mathrm{MT} & \mathrm{DT} \end{bmatrix} \tag{5.9}$$

where the rows represent non-targets and targets, and the columns represent non-detections and detections, respectively [Tarassenko *et al.* 1998]. Based on these numbers,

$$\text{Sensitivity} = \frac{\text{Detected Targets}}{\text{Targets}} = \frac{\text{DT}}{\text{DT} + \text{MT}} \tag{5.10}$$

and

$$\text{Selectivity} = \frac{\text{Detected Targets}}{\text{Detections}} = \frac{\text{DT}}{\text{DT} + \text{FD}} \tag{5.11}$$

as well as

$$\text{Performance} = \frac{1}{2}\left(\text{Sensitivity} + \text{Selectivity}\right) \tag{5.12}$$

of a classifier can be assessed. Another measure,

$$\text{Specificity} = \frac{\text{False Detections}}{\text{Non-Targets}} = \frac{\text{FD}}{\text{RN} + \text{FD}} \tag{5.13}$$

could be used instead of selectivity. However, since the number of non-targets is much larger than the number of targets in the detection problem considered in this work a high specificity may conceal a poor selectivity. Although some authors use specificity in the evaluation of the same detection problem [Kalayci and Özdamar 1995, Tarassenko *et al.* 1998] it is considered too unreliable by the author and therefore not used in this work.

The *probability of misclassification* (PMC) was estimated by the fraction of wrongly classified feature vectors. An additional indicator for the quality of the discriminants is the ratio of variance between samples and within samples or F-ratio derived from an *analysis of variance* (ANOVA) [Damon and Harvey 1987]. In the following, the F-ratio was determined for individual features and for the outputs of the discriminants $g^E(X)$, $g^F(X)$ and $g^Q(X)$.

Although the F-ratio is not influenced by linear transforms of the input variable, a non-linear transform can have a strong effect on its magnitude. Therefore, in order to make a fair comparison between discriminants the following modified form of the quadratic discriminant was also evaluated. In the modified form, the squared Mahalanobis distances are replaced by the Mahalanobis distances:

$$\overline{g}^Q(X) = [(X - \overline{X}^{(2)})^T S_2^{-1}(X - \overline{X}^{(2)})]^{1/2} - [(X - \overline{X}^{(1)})^T S_1^{-1}(X - \overline{X}^{(1)})]^{1/2} \ . \tag{5.14}$$

Thus $\overline{g}^Q(X)$ relates to the same dimension or 'unit' as $g^E(X)$ and $g^F(X)$. Apart from the bias, $g^Q(X)$ and $\overline{g}^Q(X)$ generate the same decision boundary. Since the bias is an additive term it does not affect the resulting F-ratios.

5.3.2 Simulation A: Equal Sample Sizes and Variances of Features

For simulation A samples ω_1 and ω_2 were generated with equal sizes $N_1 = N_2 = 500$ and standard deviation 1.0 and 0.5, respectively. For both samples, variances were equal between features. In addition to the discriminants, classification based on proximity to the sample means of features 1 and 2 was evaluated. This is equivalent to the applying one-dimensional feature subspaces to the Euclidian classifier. Classification results are listed in Table 5.1 and shown in Fig. 5.1. Although samples ω_1 and ω_2 are not linearly separable, over 93% sensitivity along with about 80% selectivity can be achieved with only one of the features alone. Euclidian and Fisher classifiers are almost equivalent

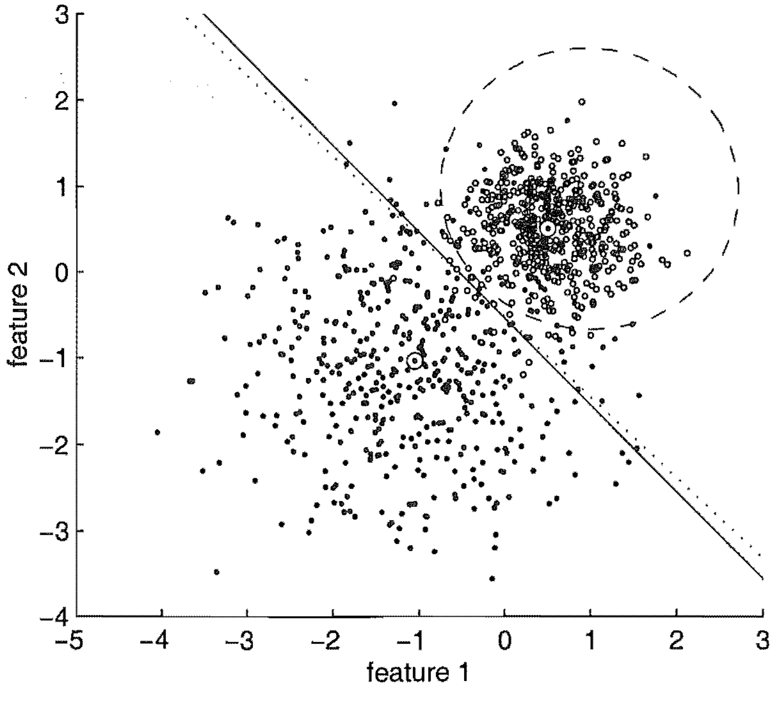

Figure 5.1 Simulation A: Gaussian distributed samples ω_1 (non-targets, black dots) and ω_2 (targets, small circles). Sample sizes were equal $N_1 = N_2 = 500$. Circles ($\odot$) show sample means. Lines show decision boundaries given by Euclidean distance classifier ($g^E(X) = 0$, solid line), Fisher's linear discriminant ($g^F(X) = 0$, dotted line) and the quadratic discriminant function ($g^Q(X) = 0$, dashed line).

in their decision boundaries and classification performance. Both achieve the highest sensitivity (98%). The quadratic classifier is more balanced between sensitivity and selectivity and achieves the highest selectivity (91%) and performance (93%). The F-ratios reflect the performance of the discriminants (for the quadratic classifier the modified version $\bar{g}^Q(X)$ needs to be considered).

Table 5.1 Classification results for simulation A: equal sample sizes and equal variances of features: F-ratio derived from ANOVA, numbers of false detections (FD), rejected non-targets (RN), detected targets (DT) and missed targets (MT); probability of misclassification (PMC), sensitivity (Sen), selectivity (Sel) and mean of sensitivity and selectivity (Per). F-ratios for both $g^Q(X)$ and $\overline{g}^Q(X)$ are listed.

Discriminant	F	FD	RN	DT	MT	PMC	Sen	Sel	Per
feature 1	913.4	118	382	469	31	14.9	93.8	79.9	86.8
feature 2	970.4	116	384	469	31	14.7	93.8	80.2	87.0
Euclidian	1839.5	72	428	490	10	8.2	98.0	87.2	92.6
Fisher	1842.1	71	429	490	10	8.1	98.0	87.3	92.7
quadratic	1136.4/2560.5	48	452	477	23	7.1	95.4	90.9	93.1

5.3.3 Simulation B: Unequal Sample Sizes and Equal Variances of Features

The sample size of ω_2 was reduced to $N_2 = 250$ while all other parameters of the samples were maintained. Classification results are listed in Table 5.2 and shown in Fig. 5.2. The decision boundary of the Euclidian classifier is not affected by the change in sample size but its sensitivity has dropped by about 10%. For the Fisher classifier, the decision boundary has moved towards the smaller sample. This is due to the the imbalance of sample sizes being compensated by the bias term in Eq. 5.4. As a result, the selectivity is maintained but the sensitivity has been considerably reduced (98% down to 82%). Here it becomes obvious how the tradeoff between sensitivity and selectivity can be controlled by changing the bias. Again, the quadratic classifier achieves the highest performance (90%).

Table 5.2 Classification results for simulation B: unequal sample sizes and equal variances of features: F-ratio derived from ANOVA, numbers of false detections (FD), rejected non-targets (RN), detected targets (DT) and missed targets (MT); probability of misclassification (PMC), sensitivity (Sen), selectivity (Sel) and mean of sensitivity and selectivity (Per).

Discriminant	F	FD	RN	DT	MT	PMC	Sen	Sel	Per
feature 1	490.7	119	381	234	16	18.0	93.6	66.3	79.9
feature 2	543.9	115	385	233	17	17.6	93.2	67.0	80.1
Euclidian	998.5	71	429	245	5	10.1	98.0	77.5	87.8
Fisher	999.2	26	474	205	45	9.5	82.0	88.7	85.4
quadratic	595.8/1419.3	44	456	237	13	7.6	94.8	84.3	89.6

5.3.4 Simulation C: Equal Sample Sizes and Unequal Variances of Features

To illustrate the adaptability of the discriminants to another situation, samples with equal sizes $N_1 = N_2 = 500$ but unequal variances in their features were generated. For sample ω_1 features 1 and 2 were generated with standard deviations of 1 and 0.5,

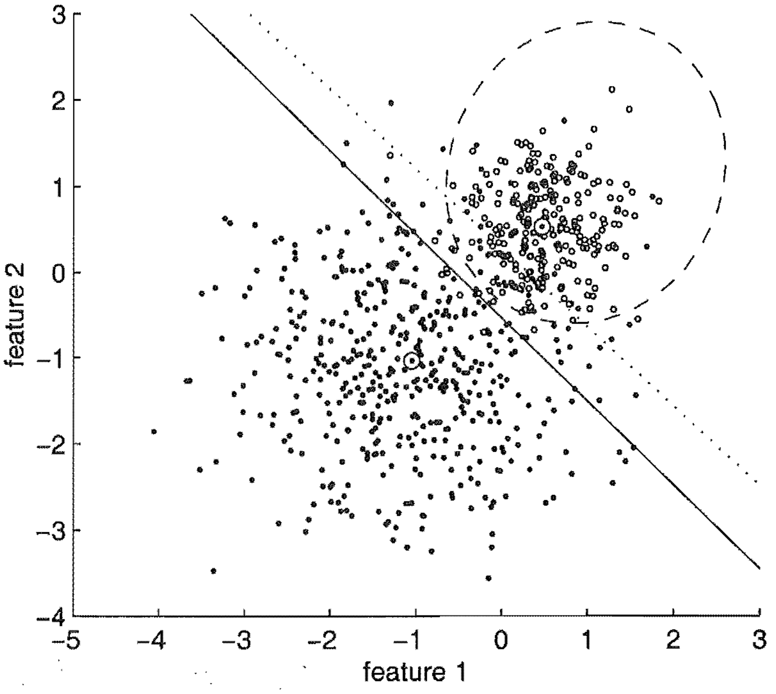

Figure 5.2 Simulation B: Gaussian distributed samples ω_1 (black dots) and ω_2 (small circles). Sample sizes were $N_1 = 500$ and $N_2 = 250$. Lines show decision boundaries given by Euclidean distance classifier (solid line), Fisher's linear discriminant (dotted line) and the quadratic discriminant function (dashed line).

respectively. For sample ω_2 both features had standard deviations of 0.5. Classification results are listed in Table 5.3 and shown in Fig. 5.3. The Fisher classifier features a decision boundary which is tilted with respect the Euclidian classifier. This gives the Fisher classifier increased sensitivity and selectivity and a performance which is close to the quadratic classifier.

Table 5.3 Classification results for simulation C: equal sample sizes, but unequal variance of features for sample ω_1: F-ratio derived from ANOVA, numbers of false detections (FD), rejected non-targets (RN), detected targets (DT) and missed targets (MT); probability of misclassification (PMC), sensitivity (Sen), selectivity (Sel) and mean of sensitivity and selectivity (Per).

Discriminant	F	FD	RN	DT	MT	PMC	Sen	Sel	Per
feature 1	913.4	118	382	469	31	14.9	93.8	79.9	86.8
feature 2	2413.1	30	470	468	32	6.2	93.6	94.0	93.8
Euclidian	2571.1	53	447	489	11	6.4	97.8	90.2	94.0
Fisher	3277.4	23	477	484	16	3.9	96.8	95.5	96.1
quadratic	2153.7/4469.0	19	481	484	16	3.5	96.8	96.2	96.5

5.4 MORE DIMENSIONS AND VARYING STATISTICS

The detection problem in medical diagnosis often involves a large sample of healthy or 'normal' cases but only a small sample with unequivocal abnormalities. To evaluate

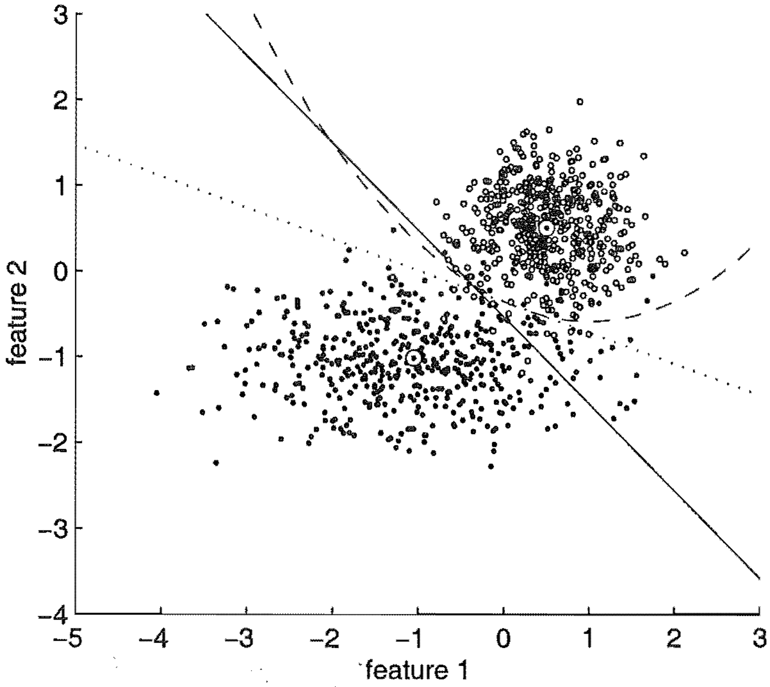

Figure 5.3 Simulation C: Gaussian distributed samples ω_1 (black dots) and ω_2 (small circles). Sample sizes were equal $N_1 = N_2 = 500$. Standard deviation of feature 2 of of sample ω_1 was reduced to 0.5. Lines show decision boundaries given by Euclidean distance classifier (solid line), Fisher's linear discriminant (dotted line) and the quadratic discriminant function (dashed line).

the mean performance and robustness of the classifiers in this situation, ω_1 and ω_2 were given sample sizes $N_1 = 2000$ and $N_2 = 200$ respectively.

Three sets of simulations were performed (D, E and F). Simulations were repeated 5000 times in each set and all performance properties were averaged. Sample ω_1 was centered around the origin. In each of the 5000 runs several properties of the samples were varied:

- the means of the features of sample ω_2 were translated to a random offset between 0.1 and 1.0
- standard deviations of all features were chosen at random between 0.1 and 1.0
- sample ω_1 was rotated at random to simulate correlated and uncorrelated features.

In simulation set D, two-dimensional Gaussian-distributed features were simulated. In simulation set E, five-dimensional Gaussian-distributed features were simulated. Finally, in simulation set F, five-dimensional non-Gaussian features were simulated.

In each run, two sets of samples with equal statistical properties were generated. One set was used for the training of the classifiers while the other one functioned as a test set. Both Fisher and quadratic classifiers were evaluated with and without bias terms to study the effect of bias correction on the overall performance. For the quadratic classifier, F-ratios were determined for both $g^Q(X)$ and $\overline{g}^Q(X)$.

5.4.1 Simulation Set D: Two-dimensional Gaussian Samples

The 2D case is comparable with the illustrated simulations in the previous section. It should be noted that the average Euclidian distance of sample means was only $d = |X^{(1)} - X^{(2)}| = \sqrt{2 * 0.55^2} = .78$. Classification results (Table 5.4) show that the

Table 5.4 Mean classification results for simulation set D: 2D Gaussian samples of grossly unequal sample size: F-ratio derived from ANOVA, numbers of false detections (FD), rejected non-targets (RN), detected targets (DT) and missed targets (MT); probability of misclassification (PMC), sensitivity (Sen), selectivity (Sel) and mean of sensitivity and selectivity (Per).

Discriminant	$\overline{F}$	$\overline{\text{FD}}$	$\overline{\text{RN}}$	$\overline{\text{DT}}$	$\overline{\text{MT}}$	$\overline{\text{PMC}}$	$\overline{\text{Sen}}$	$\overline{\text{Sel}}$	$\overline{\text{Per}}$
Euclidian	662.0	441	1559	155	45	22.1	77.5	32.6	55.1
Fisher unbiased	838.5	389	1611	153	47	19.8	76.7	37.3	57.0
Fisher	838.5	5	1995	49	151	7.1	24.7	69.1	46.9
quadratic unbsd	793.8/1282.5	416	1584	158	42	20.8	78.9	36.3	57.6
quadratic	793.8/1282.5	94	1906	128	72	7.6	64.0	59.3	61.7

unbiased classifiers outperform the biased variants. The overall mean performance is relatively poor, no classifier achieves more than 50% performance. Apparently, two features of the proposed simulation model do not provide enough information to perform classification with reasonable reliability in the situation simulated.

5.4.2 Simulation Set E: Five-dimensional Gaussian Samples

In the case of 5 dimensions, more information is available for classification of each case. In a set of five features there is a better chance of finding one feature that gives good discrimination. Furthermore, the average Euclidian distance of sample means was $d = \sqrt{5 * .55^2} = 1.23$. Because of the higher dimensionality samples are further apart than

Table 5.5 Mean classification results for simulation set E: 5D Gaussian samples of unequal sample size: F-ratio derived from ANOVA, numbers of false detections (FD), rejected non-targets (RN), detected targets (DT) and missed targets (MT); probability of misclassification (PMC), sensitivity (Sen), selectivity (Sel) and mean of sensitivity and selectivity (Per).

Discriminant	$\overline{F}$	$\overline{\text{FD}}$	$\overline{\text{RN}}$	$\overline{\text{DT}}$	$\overline{\text{MT}}$	$\overline{\text{PMC}}$	$\overline{\text{Sen}}$	$\overline{\text{Sel}}$	$\overline{\text{Per}}$
Euclidian	1218.9	255	1745	174	26	12.8	86.8	46.4	66.6
Fisher unbiased	2076.9	154	1846	168	32	8.4	84.0	62.3	73.1
Fisher	2076.9	4	1996	99	101	4.8	49.6	92.7	71.1
quadratic unbsd	1249.1/2343.9	157	1843	183	17	7.9	91.5	65.4	78.4
quadratic	1249.1/2343.9	42	1958	179	21	2.9	89.5	82.1	85.8

in simulation set D. This needs to be considered when comparing the two-dimensional simulations with the five-dimensional simulations. The classification results (Table 5.5) show superior performance for the biased classifiers. Again, the best performance is achieved by the quadratic classifier with an average of only 42 missed cases and 21 false detections.

5.4.3 Simulation Set F: Five-dimensional non-Gaussian Samples

It is important to study the effect non-normal distributions have on the performance of the classifiers. In many applications, variables do not follow the normal distributions. In the following simulations, variables were generated following either a

- χ^2-distribution with 2 degrees of freedom
- an exponential distribution ($\mu = 1$)
- or a uniform distribution in the interval [-.5 .5].

Table 5.6 Mean classification results for simulation set F: 5D non-Gaussian samples of grossly unequal sample size: F-ratio derived from ANOVA, numbers of false detections (FD), rejected non-targets (RN), detected targets (DT) and missed targets (MT); probability of misclassification (PMC), sensitivity (Sen), selectivity (Sel) and mean of sensitivity and selectivity (Per).

Discriminant	$\overline{F}$	$\overline{\text{FD}}$	$\overline{\text{RN}}$	$\overline{\text{DT}}$	$\overline{\text{MT}}$	$\overline{\text{PMC}}$	$\overline{\text{Sen}}$	$\overline{\text{Sel}}$	$\overline{\text{Per}}$
Euclidian	952.1	237	1763	153	47	12.9	76.7	46.8	61.8
Fisher unbiased	1236.5	152	1848	150	50	9.2	74.8	60.4	67.6
Fisher	1236.5	6	1994	95	105	5.0	47.6	93.2	70.4
quadratic unbsd	2108.9/3816.8	1179	821	198	2	53.7	99.2	15.4	57.3
quadratic	2108.9/3816.8	67	1933	176	24	4.1	88.1	73.3	80.7

The classification results (Table 5.6) are only slightly degraded compared to the Gaussian samples. The quadratic classifier performs exceptinally poor without the application of the bias. Again, the best performance is achieved by the biased quadratic classifier with an average of only 67 missed cases and 24 false detections.

5.4.4 Receiver Operating Characteristics

A comprehensive picture of the performance of the classifiers can be gathered from their *receiver operating characteristic* (ROC). The ROC is a function which summarizes the possible performances of an observer faced with the task of detecting a signal in noise [Egan 1975]. To obtain an ROC curve for two given samples and a given classifier, the threshold of the corresponding detection function is varied such that all false detection probabilities are covered. For each false detection probability the respective sensitivity is obtained. The area under an ROC curve equals one for a perfect detector and 0.5 for a random detector.

Mean ROC curves were obtained for simulations E and F. For a given false detection probability, detection probabilities were averaged over 5000 repetitions. Mean ROC curves were calculated for the Euclidian classifier (one feature and all features), Fisher classifier and quadratic classifier (both variants $g^Q(X)$ and $\overline{g}^Q(X)$). Fig. 5.4 shows the resulting ROC curves for five-dimensional Gaussian and non-Gaussian distributions. The ROC curves show the performance of the classifiers in a straightforward manner.

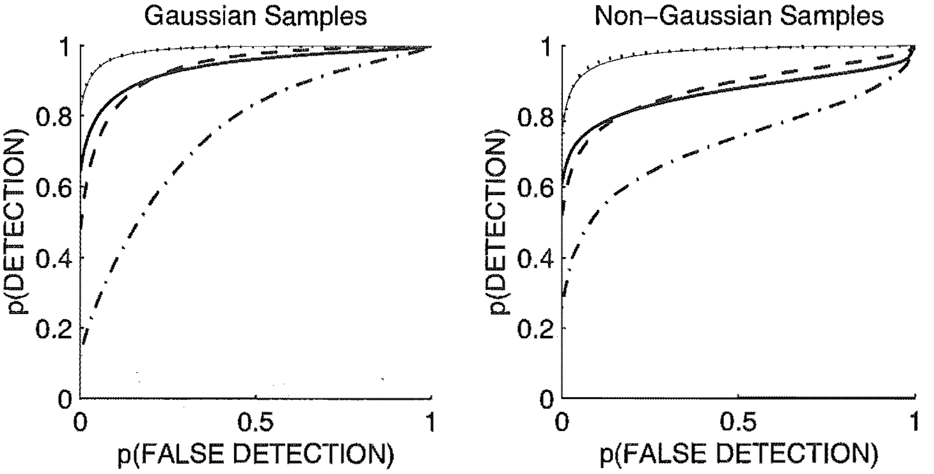

Figure 5.4 Mean ROC curves for simulations E and F: Euclidian classifier (dash dotted line: one feature, solid line: all features), Fisher classifier (dashed line) and quadratic classifier ($g^Q(X)$: dotted line, $\overline{g}^Q(X)$: thin line).

The tradeoff between sensitivity and selectivity is displayed in an obvious way that allows for comparison of classifiers without the need to choose a bias. Any point on an ROC curve can be chosen for a design by adjusting the bias.

5.5 BAYESIAN PROBABILITIES

Statistical classifiers naturally provide a binary decision in an automatic fashion. A given sample X is either classified as a member or a non-member of some class. However, if classes overlap significantly a decision may not be definite. The uncertainty of the decision cannot be reflected by the binary output. In the following, a Bayesian approach based on the class distributions *after* application of the discriminant function is presented.

Histograms were generated for the output of the quadratic classifier ($\overline{g}^Q(X)$) for the two 2D Gaussian samples with unequal sample sizes (simulation B) as shown in Fig. 5.2. The Bayesian probability of class membership $p(\omega_2|X)$ was computed according to Eq. 5.8 for all non-zero histogram bins (Fig. 5.5). Thresholds for probabilistic terms such as questionable, possible, probable, or definite, could be derived from the membership probabilities. For example, for the simulations shown, these thresholds provide an estimate for the confidence in a decision made by the classifier. However,

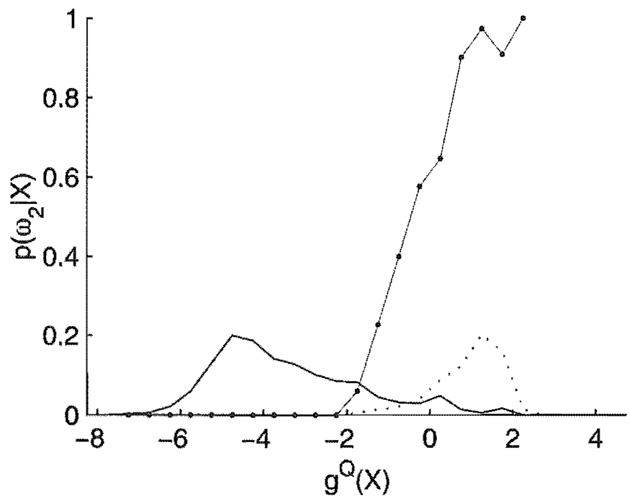

Figure 5.5 Normalized histograms of the output of the quadratic classifier ($\overline{g}^Q(X)$) for Gaussian samples ω_1 and ω_2 with unequal sample sizes ($N_1 = 500$, $N_2 = 250$). The Bayesian probability of class membership $p(\omega_2|X)$ is indicated by the thin line and dots for relevant bins.

the histogram bins have to be found empirically since no general method is available for their automatic computation.

5.6 SUMMARY

Linear and quadratic statistical classifiers were presented and their ability to discriminate between samples from two classes was tested. To simulate a variety of situations test samples included

- Gaussian and non-Gaussian distributions,
- equal and unequal samples sizes,
- equal and unequal variances in individual variables, and
- correlated and uncorrelated variables.

Several performance indicators were presented and applied to the classifiers in repeated tests. The mean of all performance indicators over 5000 repetitions shows the robustness and weaknesses of the individual classifiers in various situations. Both Fisher linear and quadratic classifiers show robust performance even for non-Gaussian distributed samples.

A Bayesian approach to obtain the confidence of a classification with respect to a given threshold was presented. This approach can be used to validate the decision made by a statistical classifier.

Chapter 6

SPATIAL ANALYSIS OF THE ELECTROENCEPHALOGRAM

6.1 INTRODUCTION

The spatial distribution of electrical activity seen in the channels of the EEG provides important clues for the classification of artifacts and pathological events. Several rules can be applied to an electrode chain in a bipolar recording to identify focal patterns and provide a coarse indication of focus localization. A more sophisticated approach is to model the electric field caused by a focal source within the brain, which is commonly modelled as a *dipole*. The six degrees of freedom (location, orientation and strength) of a dipole can be optimized to match some observed spatial distribution given a model of the head and its conductivity distributions. However, it is important to note that events can arise from several distinct and/or distributed sources making accurate localization impossible.

Importantly, the aim of this work is to extract a feature from the measured potential distribution that helps distinguish epileptiform activity from normal background. The clinical literature suggests that the polarity of the surface potential is the most indicative feature of EDs [Duffy *et al.* 1989, Binnie 1993]. Thus, an attempt is made to estimate both location and orientation of the source.

In the following sections the three important spaces relevant for the spatial analysis of the EEG are introduced. Forward models are presented and applied to illustrate the relationships between the three spaces: the surface potential resulting from a localized source, the vector field in the source volume that relates to a single channel of the EEG, and the 3D sensitivity distribution for a given montage. Eventually, an approach to source localization and orientation estimation is presented and tested in simulations.

6.2 RELEVANT SOURCE AND SIGNAL SPACES

Several spaces that play an important role in the generation and analysis of the EEG are outlined in the following subsections. Understanding their interrelationships is important for the interpretation of events in the EEG.

6.2.1 Electrode Space and Absolute Reference

Electrode space can be defined as a multidimensional signal space with one dimension for each electrode. Electrodes sample the surface potential of the head. Ideally, the potential would be measured with respect to some absolute zero reference potential. However, no such absolute zero potential exists and only potential differences between pairs of electrodes can be recorded. In some cases one electrode, such as on an earlobe (A1 or A2) or vertex (Cz), may be used as a common reference for several other electrodes. Movements or changes in the conductivity properties of a single electrode cause *electrode artifacts* that are localized in electrode space but will usually affect several channels. Artifacts on a reference electrode may contaminate the entire recording.

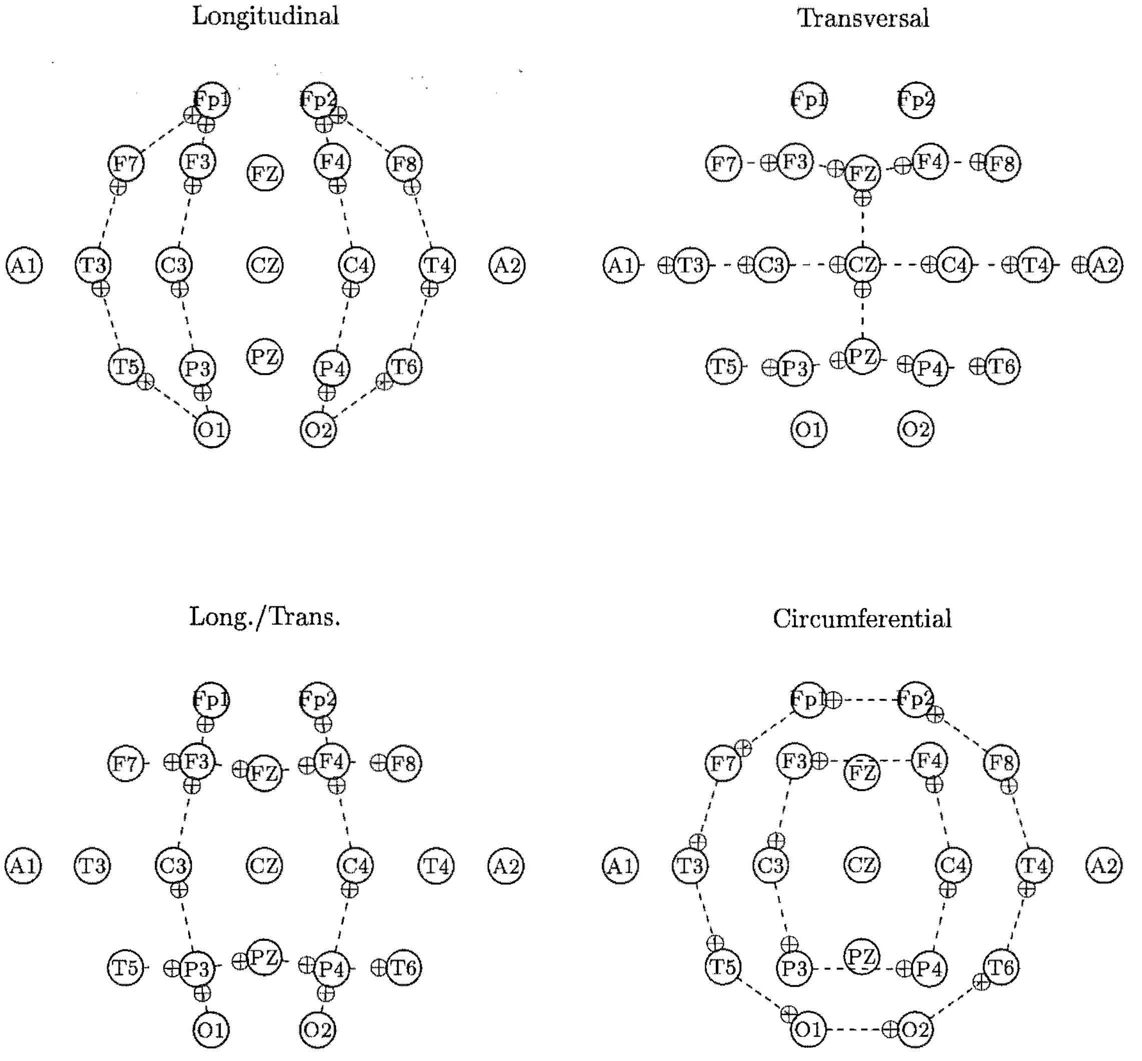

Figure 6.1 Bipolar montages used in 16-channel clinical recordings. The symbol ⊕ indicates the positive electrode of each channel while the other electrode serves as reference.

6.2.2 Channel Space and Montages

The mapping of electrode pairs to channels is defined by the *montage* of a recording. A *referential montage* requires an assumed absolute reference such as the ipsilateral earlobe potential. The average of all channels is also sometimes used as a common reference. In contrast, *bipolar montages* do not have a common reference. The bipolar montages used in the clinical recordings analyzed in this thesis are shown in Fig. 6.1. Bipolar montages can give a more detailed picture of more superficial focal events whereas referential montages can give a more sensitive representation of deeper sources [Katznelson 1981]. Sensitivity patterns will be illustrated in detail in the following sections. Each channel has a specific pattern of sensitivity to a certain region of the source space called *lead field.* Thus, a focal source may be picked up in one channel but may be obscured in another. The term 'lead field' was coined in *electrocardiography* (ECG), where a 'lead' is the voltage measured by a particular combination of electrodes [Goldman 1973].

The channel space can be defined as a multidimensional signal space with one dimension for each channel. It is a subspace of the electrode space. A montage corresponds to a non-invertible transfer matrix from electrode space to channel space.

6.2.3 Source Space and Electrostatic Models

The source space can be defined as a multidimensional current density distribution inside the physical volume of the head. At any particular location the current density is specified by the three-dimensional vector that represents the strength and the orientation of a local current source. It is generally assumed that the volume conduction is instantaneous within the relevant time resolution and that resistivity is constant with respect to the frequencies found in the EEG, such that an electrostatic model is sufficient as opposed to an electrodynamic model [Nunez 1981]. Since current fields and potentials of several sources do not interact during volume conduction, they can be added linearly.

A local current source is modelled by a *dipole* or *equivalent current dipole* (ECD). A dipole represents the average of some locally-distributed synchronous activity which can be resolved within the temporal and spatial resolution limits of the EEG.

Several *forward models* have been suggested that allow for the calculation of the surface potential for a given dipole. Some models assume one or more concentric spherical shells of homogenous conductivity, representing the brain, *cerebrospinal fluid* (CSF), skull or scalp. More advanced models use finite element techniques to closely render these compartments according the anatomy of individual patients. Several forward models will be discussed in the following section.

6.3 FORWARD SOLUTION FOR FOUR-SHELL SPHERICAL MODEL

The electrostatic properties of the head can be modelled as a set of four concentric shells of homogenous and isotropic conductivities that represent scalp, skull, subarachnoid space (containing CSF), and the brain. There are analytic expressions for the

Table 6.1 Spherical head model parameters. The model proposed by Rush and Driscoll consists of three shells only.

Model	Conductivities ($1/\Omega m$)				Boundary radii		
	Scalp σ_4	Skull σ_3	CSF σ_2	Brain σ_1	Skull d	CSF c	Brain b
Rush and Driscoll	0.33	0.0041		0.33	0.9200		0.8700
Cuffin and Cohen	0.33	0.0042	1.0	0.33	0.9659	0.9295	0.8977
Stok	0.33	0.0042	1.0	0.33	0.9467	0.8667	0.8400

surface potential generated by a single source inside these volume conductor models. For spherical head models, a number of conductivities and shell radii are assumed. Several three- and four-shell head models with various conductivities and radii have been proposed, some of which are listed in Table 6.1 [Rush and Driscoll 1968, Cuffin and Cohen 1979, Stok 1986, Stok 1987]. The large difference in conductivity of the skull with respect to the other compartments ($\sigma_{brain}/\sigma_{skull} \simeq 80$) causes most of the blur in the surface potential fields [Berg and Scherg 1994]. The model proposed by Stok [1986] is shown in Fig. 6.2. With the conductivities and eccentricities from Table 6.1 one can

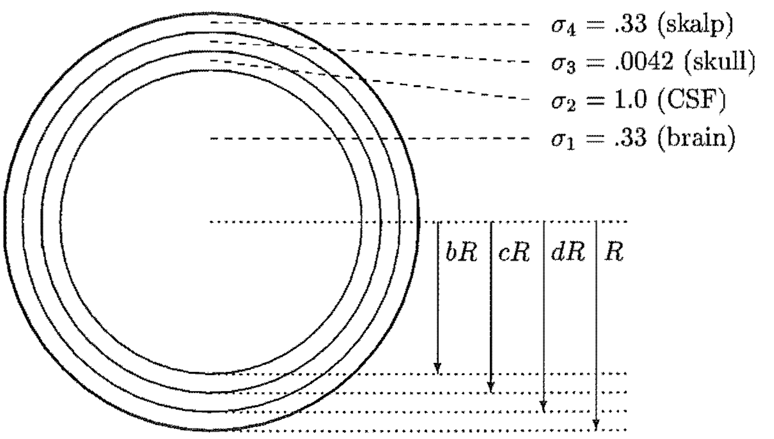

Figure 6.2 Spherical head model with shell radii $b = .84$ (brain/CSF), $c = .8667$ (CSF/skull) and $d = .9467$ (skull/scalp). Shell radii are not drawn at their original scale.

calculate the potential v at some surface location $\mathbf{s} = (s_x, s_y, s_z)$ for a dipole at location $\mathbf{r} = (r_x, r_y, r_z)$ with current dipole moment $\mathbf{m} = (m_x, m_y, m_z)$ according to

$$v = \frac{1}{4\pi\sigma_4 R^2} \sum_{n=1}^{\infty} \Gamma^{\text{scalp}}(n)\, f^{n-1} \mathbf{m} \cdot \left[\mathbf{r}_0 P_n(\cos\theta) + \mathbf{t}_0 \frac{P_n^1(\cos\theta)}{n} \right] \tag{6.1}$$

with head radius $R = |\mathbf{s}|$, dipole eccentricity $f = |\mathbf{r}|/R$, radial unit vector $\mathbf{r_0}$, tangential unit vector $\mathbf{t_0}$, angle θ between $\mathbf{r}$ and $\mathbf{s}$ ($\cos\theta = \mathbf{r}\cdot\mathbf{s}$), Legendre polynomials $P_n(\cos\theta)$ and associated Legendre polynomials $P_n^1(\cos\theta)$ of degree n. The weights Γ^{scalp} are

$$\Gamma^{\text{scalp}}(n) = \frac{(2n+1)^4(cd)^{2n+1}}{\Gamma(n)} \tag{6.2}$$

with

$$\begin{aligned}
\Gamma(n) = d^{2n+1} & \left[b^{2n+1}n\left(\frac{\sigma_1}{\sigma_2}-1\right)\left(\frac{\sigma_2}{\sigma_3}-1\right)(n+1) + c^{2n+1}\left(\frac{\sigma_1}{\sigma_2}n+n+1\right)\left(\frac{\sigma_2}{\sigma_3}n+n+1\right)\right] \\
& \cdot\left[\left(\frac{\sigma_3}{\sigma_4}n+n+1\right)+(n+1)\left(\frac{\sigma_3}{\sigma_4}-1\right)d^{2n+1}\right] \\
+(n+1)c^{2n+1} & \left[b^{2n+1}\left(\frac{\sigma_1}{\sigma_2}-1\right)\left(\frac{\sigma_2}{\sigma_3}n+\frac{\sigma_2}{\sigma_3}+n\right)+c^{2n+1}\left(\frac{\sigma_1}{\sigma_2}n+n+1\right)\left(\frac{\sigma_2}{\sigma_3}-1\right)\right] \\
& \cdot\left[n\left(\frac{\sigma_3}{\sigma_4}-1\right)+\left(\frac{\sigma_3}{\sigma_4}n+\frac{\sigma_3}{\sigma_4}+n\right)d^{2n+1}\right]
\end{aligned} \tag{6.3}$$

[Salu *et al.* 1990, Sun 1997]. To find the surface potential of a single dipole at one location the extensive calculations have to be repeated for 60-100 terms to approximate the infinite sum in Eq. 6.1.

To speed up calculations, Nunez [1981] suggested estimating the magnitude of the surface potential with a simple homogenous-conductor model and subsequently multiplying the result by a correction for inhomogeneity. Several fast algorithms for the forward calculations have been proposed. Berg and Scherg [1994] approximated the four-shell model by three separate dipoles of equal orientation in a sphere of homogenous conductivity. They determined the weights and relative eccentricities of the approximation dipoles numerically. Their approximation is 33 times faster than the standard method (empirical residual variance $< 1\cdot 10^{-4}$). Sun [1997] calculated the first three elements ($n = 1, 2, 3$) of the term $\Gamma^{\text{scalp}}(n)f^{n-1}$ in Eq. 6.1 explicitly and approximated the rest ($n > 3$) with a polynomial. He presents error bounds in analytical form and his approximation is 50-100% faster than the one proposed by Berg and Scherg [1994].

6.3.1 Forward Solution for Radial Dipoles

The four-shell model proposed by Stok [1986] was used with the algorithm given by Sun [1997] to estimate the surface potentials of a radial dipole at various eccentricities. Surface potentials are plotted radially with respect to the outer surface (darkest solid line) in a cross section (Fig. 6.3) presented in a manner similar to that of Junghöfer *et al.* [1997]. The shallow source produces the sharpest peak in the surface potential while the deep source leads to a more even distribution. The peak absolute surface

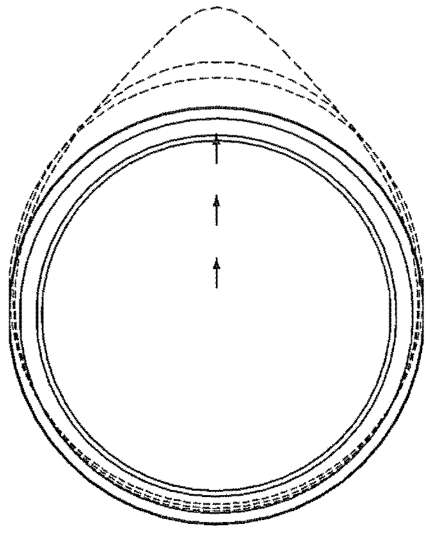

Figure 6.3 Relative potentials of radial dipoles at eccentricities 0.8, 0.5 and 0.2, respectively. Positive potentials are shown outside the scalp surface.

potential of the intermediate source is about half that of the shallow source.

6.3.2 Forward Solution for Tangential Dipoles

The cross section of surface potentials generated by tangential sources are shown in Fig. 6.4. Tangential sources generally produce lower amplitude surface potentials than radial sources, but shallow tangential sources generate significant gradients. Most of the depth-related difference in the potential distribution is found in the 90 degree sector above the sources.

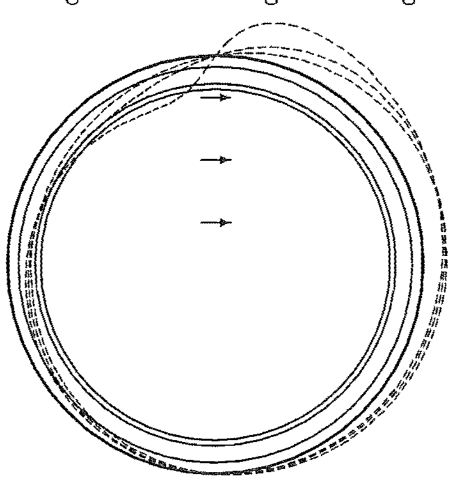

Figure 6.4 Relative potentials of tangential dipoles at eccentricities 0.8, 0.5 and 0.2, respectively. Positive potentials are shown outside the scalp surface. The same scale as in the previous figure is used.

6.3.3 Implications of Forward Solutions

For both radial and tangential sources, the depth influences the scale of the potential pattern at the surface. In other words, the surface potential pattern of deep sources has lower amplitude than that of shallow sources. Conversely, the estimation of source eccentricity could be compared to the estimation of the scale of a spatial low-pass (blur) filter.

In bipolar montages, channels measure the gradient of the potential distribution. For a tangential source the gradient is maximal above the source. In the case of a radial source it is maximal along a closed curve circumscribing the source.

For the average reference montage, the average of all channels is used as a reference but electrodes cover the upper half of the head only. Since dipole potentials for both radial and tangential sources extend over the entire surface of the head the average reference can be biased.

6.3.4 Boundary Element Models

More detailed models of the homogenous regions with respect to their conductivity can be derived from a T1-weighted *magnetic resonance image* (MRI) of the head. The reconstructed surfaces between homogenous regions are used in *boundary element method* (BEM) models that lead to more accurate solutions. Inner and outer surfaces of the BEM models are constructed with triangular elements. The triangular elements sample a 2D surface in 3D space. Ferguson and Stroink [1997] show how the density of the surface tessellation of a spherical surface influences the accuracy of the BEM model. They compared several approaches to approximate the surface integral of the BEM model with the discrete surface elements. Fuchs *et al.* [1998] proposed a way to refine the mesh size of triangular surface elements for general surfaces that leads to a markedly more accurate forward solution. Both the three-shell spherical model and a three-compartment BEM model based on their new approach were used to fit dipoles to nine time samples of an from a focal ED recorded with 28 electrodes. The BEM model resulted in less spatial spread of the nine resulting dipoles. They concluded that the variance between measured and fitted field is only weakly correlated to the dipole mislocalization. Thus a good fit in the EEG potentials could be achieved with an oversimplified model though the dipole location could be several centimetres away from the true position of the source [Fuchs *et al.* 1998].

In this work a spherical model is used since a general spatial model is required and data sets representing the patient's individual anatomy (CT or MRI scans) were available in a few cases only.

6.4 LEAD FIELDS

The lead field is the electric current field that would result from injecting a unit current into a volume conductor with a pair of surface electrodes. According to the reciprocity theorem of Helmholtz, the distribution of this current field is proportional to the sensitivity of the lead. The lead field defines the sensitivity of any pair of electrodes and any bipolar channel. Malmivuo *et al.* [1997] calculated the *half-sensitivity volume* for several electrode separations to compare EEG and MEG sensitivity patterns. For a pair of electrodes 20^o apart, the half-sensitivity volume ends about 8 mm under the cortical surface and the sensitivity drops to about 7% at the centre of the head relative to the maximum sensitivity. For electrodes located 60^o and 180^o apart, they report a relative sensitivity of 20% and 35%, respectively, at the centre of the head.

6.4.1 Direction

Electrodes in the bipolar montages used in this work are placed about 36^o apart. The direction of the lead field current flow for a bipolar channel is illustrated in Fig. 6.5. Additionally, the case of the widest possible electrode separation (180^o) is shown. The

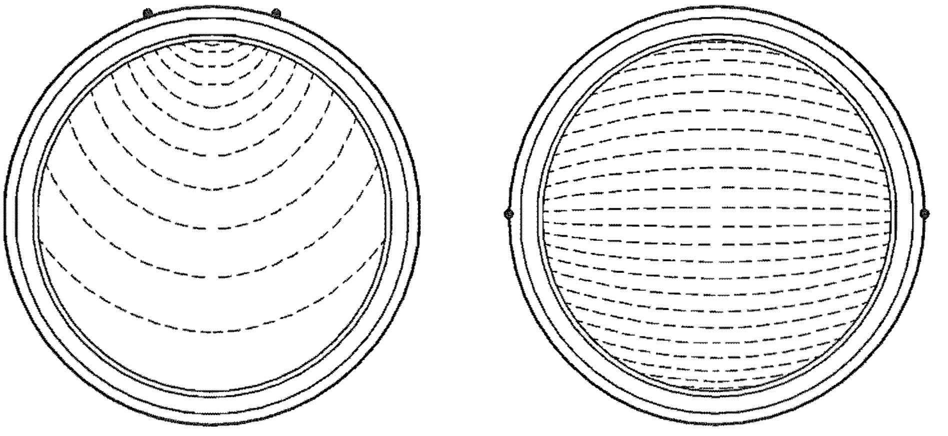

Figure 6.5 Directions of lead field current flow for electrode pairs placed 36 degrees (left) and 180 degrees (right) apart. Electrode locations are indicated by black dots.

current flow is perpendicular to the surface underneath the electrodes where the current enters or leaves the volume conductor. According to the reciprocity theorem a channel is insensitive to a current source perpendicular to the flow field. Conversely, a channel is most sensitive to sources aligned with the flow field. The illustrations indicate that bipolar channels are most sensitive to radial sources underneath the electrodes and tangential sources half way between electrodes.

6.4.2 Channel-Specific Sensitivities

The isosensitivity surfaces on which a current source of 10 nAm would generate potential differences of a range of small voltages in the electrode pairs is shown in Fig. 6.6. The larger the electrode separation, the larger is the sensitive volume covered by the

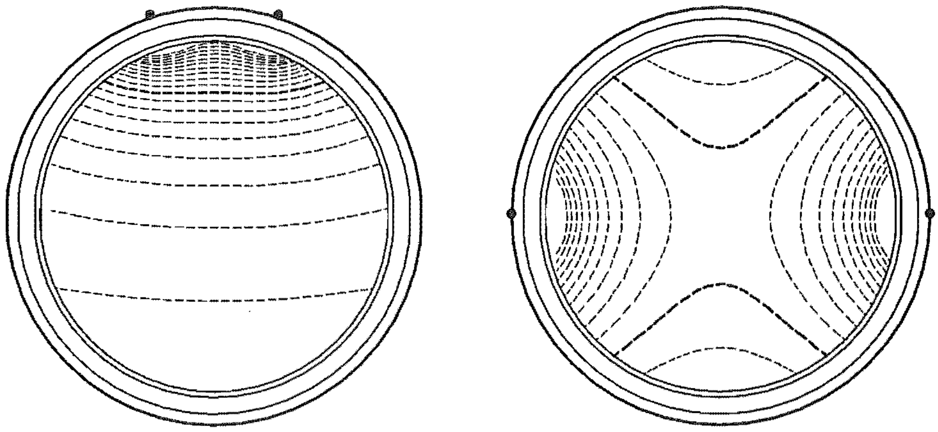

Figure 6.6 Isosensitivity surfaces of lead fields for electrode pairs as shown in the previous figure. The thick dashed line indicates the surface where a current source of 10 nAm aligned in parallel to the lead field would generate a potential difference of 1 μV in the electrode pair. Thin dashed lines closer to the electrodes show the surfaces for $0.9, 0.8 \ldots 0.1\,\mu$V, respectively. Thin dashed lines further apart from the electrodes show the surfaces for $1.1, 1.2 \ldots 1.8\,\mu$V, respectively.

channel. Since sensitivity is highest underneath the electrodes where the current flow is radial, the EEG is generally more sensitive to radially-oriented sources.

From these diagrams one can deduce that a channel derived from widely spaced electrodes is more sensitive to deep sources than a channel from closely spaced electrodes. Channels with large inter-electrode distances are found in referential montages only.

6.5 INVERSE PROBLEM

The attempt to localize current sources that generate a specific measurement in channel space is an *inverse problem*. As for virtually all inverse problems, there is no unique solution and a straightforward estimate is severely impaired by measurement noise. However, there are meaningful solutions if it can be assumed that only a single dipolar current source is involved. At least three distinct ways to approach this problem have been proposed:

- optimize the location and orientation parameters of a single dipole or a small number of dipoles,
- estimate the current source density at the cortical surface or at the scalp, or
- estimate the current source density within the 3D brain volume.

The case of a single source is described in the following subsection and an approach to estimate the contribution of a finite set current sources distributed in the 3D volume is proposed later in this section.

6.5.1 Dipole Fitting

For a given dipole location $\mathbf{r}$ the forward solutions for the three orthogonal orientations result in three scalar N-vectors in channel space $\mathbf{v_r}(\mathbf{r})$, $\mathbf{v_s}(\mathbf{r})$ and $\mathbf{v_t}(\mathbf{r})$, for the radial and both tangential orientations, respectively (N represents the number of channels; it is assumed that $N > 3$). The vector $\mathbf{v_r}$ represents a radially oriented dipole while $\mathbf{v_s}$ and $\mathbf{v_t}$ represent tangentially oriented dipoles. The two tangential dipoles are chosen to be orthogonal to each other and with respect to the radial dipole. The three orientation-specific vectors can be combined to the $N \times 3$ matrix

$$\mathbf{K}(\mathbf{r}) = [\mathbf{v_r}(\mathbf{r}), \mathbf{v_s}(\mathbf{r}), \mathbf{v_t}(\mathbf{r})] \ . \tag{6.4}$$

The matrix $\mathbf{K}(\mathbf{r})$ represents the entire relationship of source location $\mathbf{r}$ to the channel space or, alternatively, it defines the lead fields at location $\mathbf{r}$. Thus, the forward solution for a given source location $\mathbf{r}$ and current density (or dipole moment) $\mathbf{j}$ is

$$\mathbf{v}_{\text{dipole}} = \mathbf{K}(\mathbf{r})\,\mathbf{j} \tag{6.5}$$

where

$$\mathbf{j} = \begin{pmatrix} j_r \\ j_s \\ j_t \end{pmatrix} \ . \tag{6.6}$$

Since orthogonality cannot be assumed for $\mathbf{v_r}$, $\mathbf{v_s}$ and $\mathbf{v_t}$ in the N-dimensional channel space, an estimate of the inverse is obtained with

$$\hat{\mathbf{j}} = \mathbf{K}(\mathbf{r})^{+}\mathbf{v}_{\text{obs}} \tag{6.7}$$

where

$$\mathbf{K}^{+} = [\mathbf{K}^T\mathbf{K}]^{-1}\mathbf{K}^T \tag{6.8}$$

is the Moore-Penrose pseudoinverse of matrix $\mathbf{K}$ and $\mathbf{K}^T$ is the transpose of $\mathbf{K}$.

In order to fit a dipole to a set of observed potentials, location $\mathbf{r}$ is varied to optimize the goodness-of-fit with the dipole potentials. This is commonly done with an optimization method such as the simplex method [Hara *et al.* 1999]. The goodness-of-fit can be improved by adding further dipoles to the model. However, prior assumptions have to be made about the appropriate number of dipoles for a given measurement

[Berg and Scherg 1994].

The goodness-of-fit between observed potentials $\mathbf{v}_{\text{obs}}$ and potentials generated by the dipole $\mathbf{v}_{\text{dipole}}$ can be measured by the *dipolarity*

$$D = \sqrt{1 - \left(|\mathbf{v}_{\text{obs}} - \mathbf{v}_{\text{dipole}}|^2 / |\mathbf{v}_{\text{obs}}|^2\right)} \tag{6.9}$$

as suggested by Hara *et al.* [1999] or, similarly, by the *percent root-mean-square difference* (PRD)

$$r = \left(\frac{|\mathbf{v}_{\text{obs}} - \mathbf{v}_{\text{dipole}}|^2}{|\mathbf{v}_{\text{obs}}|^2} \right)^{\frac{1}{2}} \times 100\% \ . \tag{6.10}$$

Obviously, the PRD is minimal if the dipolarity is maximal. In this work the PRD is used since it provides a larger contrast for small differences.

6.5.1.1 Spatial Resolution

Mosher *et al.* [1993] investigated the localization error for a single dipole (strength: $|\mathbf{j}| = 10\,\text{nAm}$, noise: $\sigma = 0.4\,\mu\text{V}$ RMS) with referential montages including 21, 37 and 127 electrodes in simulations. They reported a localization error of 2 to 3.5 cm for the standard 10-20 electrode configuration. Even shallow sources which generate a stronger surface potential cannot be localized at higher accuracy since they are picked up by a single electrode only.

6.5.2 Estimation of Distributed Activity

Some approaches do not assume a single or a finite number of dipoles but model the sources as a continuous distribution on a 2D surface such as the cortex or the scalp or in the 3D cortical volume.

The surface Laplacian of the scalp potential or *current source density* (CSD) represents the sources and sinks of the generating currents. The CSD has been suggested as a means to improve the spatial resolution of the EEG. Implementations such as simple centre-surround filters and more advanced ones based on the expansion of the surface potential with spherical splines have been proposed. The CSD is very sensitive to noise and a dense electrode array is required to obtain a meaningful approximation [Perrin *et al.* 1987, Perrin *et al.* 1989].

Junghöfer *et al.* [1997] point out that the term Γ^{scalp} in the forward model (Eq. 6.1) of a radial dipole can be considered as a generalization of the spherical spline expansion given by Perrin *et al.* [1989]. Junghöfer *et al.* [1997] suggest modelling the continuous surface potential by a set of radial dipoles of fixed eccentricity under each electrode. Their approach allows for the estimation of both the scalp CSD and the cortical surface CSD while the eccentricity enables one to control the smoothness of the solution. The spherical spline expansion was also used by Edlinger *et al.* [1998] to estimate the "analytically deblurred" potential at the cortical surface.

Sidman *et al.* [1992] simulated a finite number of radial dipoles ($N = 160$) on a hemispherical surface of a given radius (r_T) within a spherical volume conductor of homogenous conductivity. Given a set of potential measurements at the outer surface, they estimated the potentials on a second hemispherical surface (radius $r_I > r_T$) using a *singular value decomposition* (SVD) truncated minimum norm solution of the inverse problem in a procedure they called *cortical imaging technique* (CIT). They illustrate how the estimated potential maps of several epileptiform discharges gain more detail as more eigenvalues are included in the SVD decomposition. The truncation index was chosen according to the estimated SNR of the recordings. Wang and He [1998] amended the CIT approach by applying the three-shell spherical head model proposed by Rush and Driscoll [1969].

Pascual Marqui *et al.* [1994] introduced a method called *low resolution electromagnetic tomography* (LORETA). A regular cubic grid of dipoles (arbitrary orientations) is used to estimate the current source distribution for a given EEG or MEG measurement in the 3D volume of the cortex. They minimize the 3D discrete Laplacian of the resulting current densities to maximize the smoothness of their solution. A comparison to similar methods can be found in Pascual Marqui [1999].

6.6 GEODESIC DIPOLE MODEL

In the following, a new approach to source modelling is presented. It is based on a regular grid of dipoles which are treated independent from each other. Only the local inverse solution that estimates the orientation and the goodness-of-fit of each dipole is used.

Referential montages that cover the head are commonly used in EEG recordings when source localization is the goal. If the data set to be analyzed is present in bipolar montages, techniques such as LORETA (see above) cannot be applied because they rely heavily on referential montages with regular electrode distributions [Pascual Marqui *et al.* 1994].

In the approach presented in the following subsections a fixed set of dipoles was generated to evenly cover the source space and render the lead fields. The spacing between dipoles was approximately 1.4 cm which corresponds to about half the resolution limit found by Mosher *et al.* [1993] for the 21 electrodes of the 10-20 system.

6.6.1 Geodesic Sampling Grid

Various approaches to sampling a hemispherical volume are possible. A cubic grid was used in the LORETA approach [Pascual Marqui *et al.* 1994], but this does not sample a spherical surface evenly. The ideal regular sampling in the sense of dividing the volume regularly onto equal-sized blocks would be a tetrahedral grid. Unfortunately, tetrahedra cannot be refined to similar smaller tetrahedra [Ruprecht and Müller 1998]. However, the triangular surface elements on an icosahedron can be refined to near similar smaller surface elements that approximate the spherical surface in what is known as *geodesic dome* or *n-frequency icosahedron ($n\nu$ icosa)* [Kenner 1976]. The geodesic geometry was suggested for the placement of a high-density electrode net by Tucker [1993]. To fill the spherical volume, several n-frequency icosahedra can be placed concentrically to obtain a grid that is near regular within each surface. This is the approach proposed by the author.

The radius r of each n-frequency icosahedron is chosen to be equal to the refinement frequency $r = n$. The grid resulting from the nodes of the icosahedra was truncated

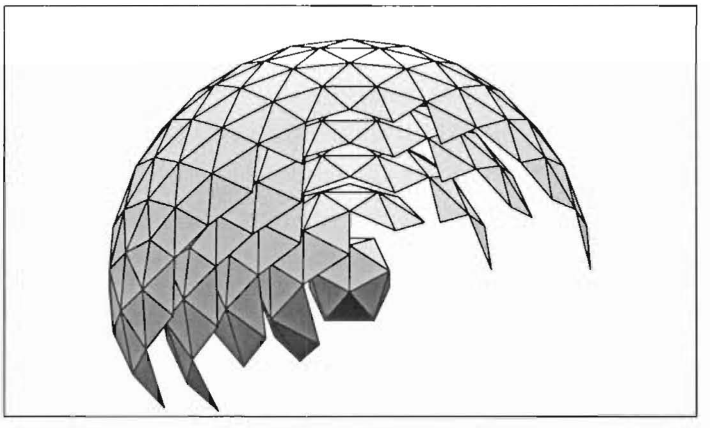

Figure 6.7 The geodesic grid defines 370 dipole locations on five hemispherical surfaces. Each surface corresponds to an n-frequency icosahedron. The size of the triangles is retained between surfaces, the distance between surfaces approximates the length of the sides of the triangles. Outer surfaces are shown incomplete to expose the inner ones.

such that only locations within the upper hemisphere were retained. Thus, $M = 370$ nodes of the grid were obtained as illustrated in Fig. 6.7 (with triangulations of the individual spheres). Each icosahedron can be projected onto a plane by an *azimuthal projection*, giving the full picture of all nodes and triangles involved (Fig. 6.8).

6.6.2 Montage-Specific Sensitivities

Forward and inverse solutions $\mathbf{K}_i$ and $\mathbf{K}_i^+$ for each node in the geodesic grid $i = 1 \ldots M$ were calculated. The *root-mean square* (RMS) of voltages observed at all channels $|\mathbf{v}_{\text{obs}}|$ in response to a unit source at some location and orientation in the head model can be

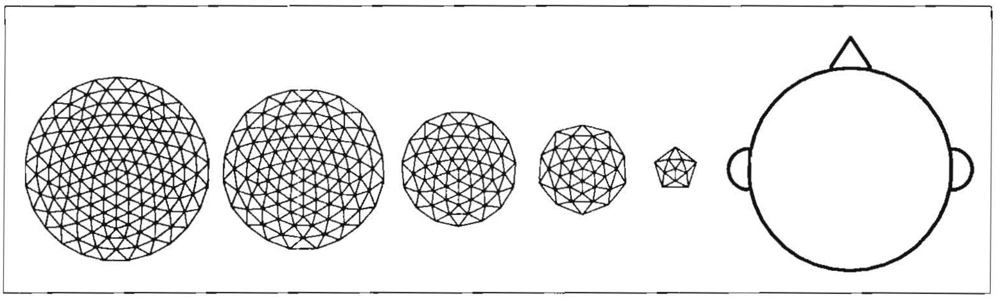

Figure 6.8 Azimuthal projection of the five n-frequency icosahedra. The centre of each disk corresponds to the location of the Cz electrode, the top to nasion and the bottom to the inion, respectively.

used as an indicator of the sensitivity achievable with a given montage. To illustrate the sensitivity patterns for bipolar and average reference montages, each of the 370 dipoles on the geodesic grid was evaluated for each montage.

6.6.2.1 Minimum Response Orientation

The strength of each dipole was fixed to a moment of 10 nAm. Dipole orientations were chosen to cause the smallest possible $|\mathbf{v}_{obs}|$. This corresponds to orientations 'most perpendicular' to the lead fields of all channels. The resulting $|\mathbf{v}_{obs}|$ values were mapped for all dipole locations and all montages in Fig. 6.9. The sensitivity patterns render the surface area covered by the electrode chains. The circumferential montage is less sensitive in the prefrontal and occipital areas since electrodes Fz and Pz are not used. Mean sensitivities within the five shells are listed in Table 6.2 for all montages. The average reference montage shows the highest sensitivity for all eccentricities. For

Table 6.2 Mean $|\mathbf{v}_{obs}|$ in μV for the five shells of the geodesic grid. Dipole orientations were chosen to minimize $|\mathbf{v}_{obs}|$. Listed are the mean responses for each shell.

Montage	Eccentricity				
	7.0 cm	5.6 cm	4.2 cm	2.8 cm	1.4 cm
Longitudinal	0.8165	0.6983	0.6474	0.5744	0.5553
Transversal	0.9088	0.7737	0.7035	0.5949	0.5610
Long./Trans.	1.0487	0.9313	0.8418	0.7360	0.6923
Circumferential	0.6108	0.4560	0.3796	0.2912	0.2600
Average Reference	1.3561	1.1812	1.0976	0.9765	0.9387

deep dipoles (eccentricity 1.4 cm) the longitudinal/transversal is the best performing bipolar montage. In this montage the earlobe reference electrodes are used (Fig. 6.1, page 84).

6.6.2.2 Radial Orientation

The field generated by radial dipoles alone is evaluated in Fig. 6.10. The sensitivity

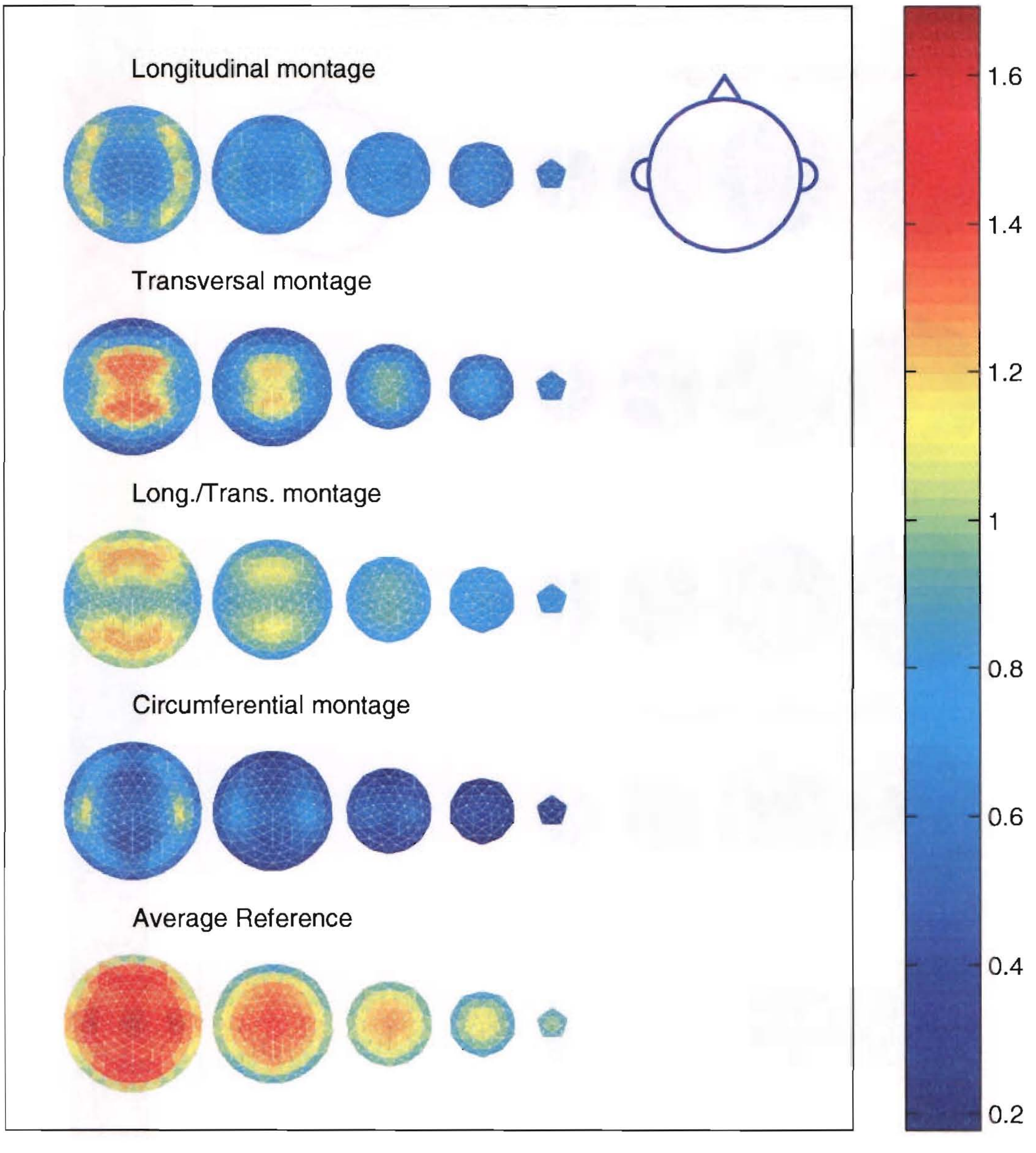

Figure 6.9 Sensitivity patterns of montages. $|\mathbf{v}_{\mathrm{obs}}|$ in μV in response to the electric fields caused by each of the 370 dipoles in the model. Dipole strengths were set to 10 nAm and dipole orientations were chosen to minimize $|\mathbf{v}_{\mathrm{obs}}|$.

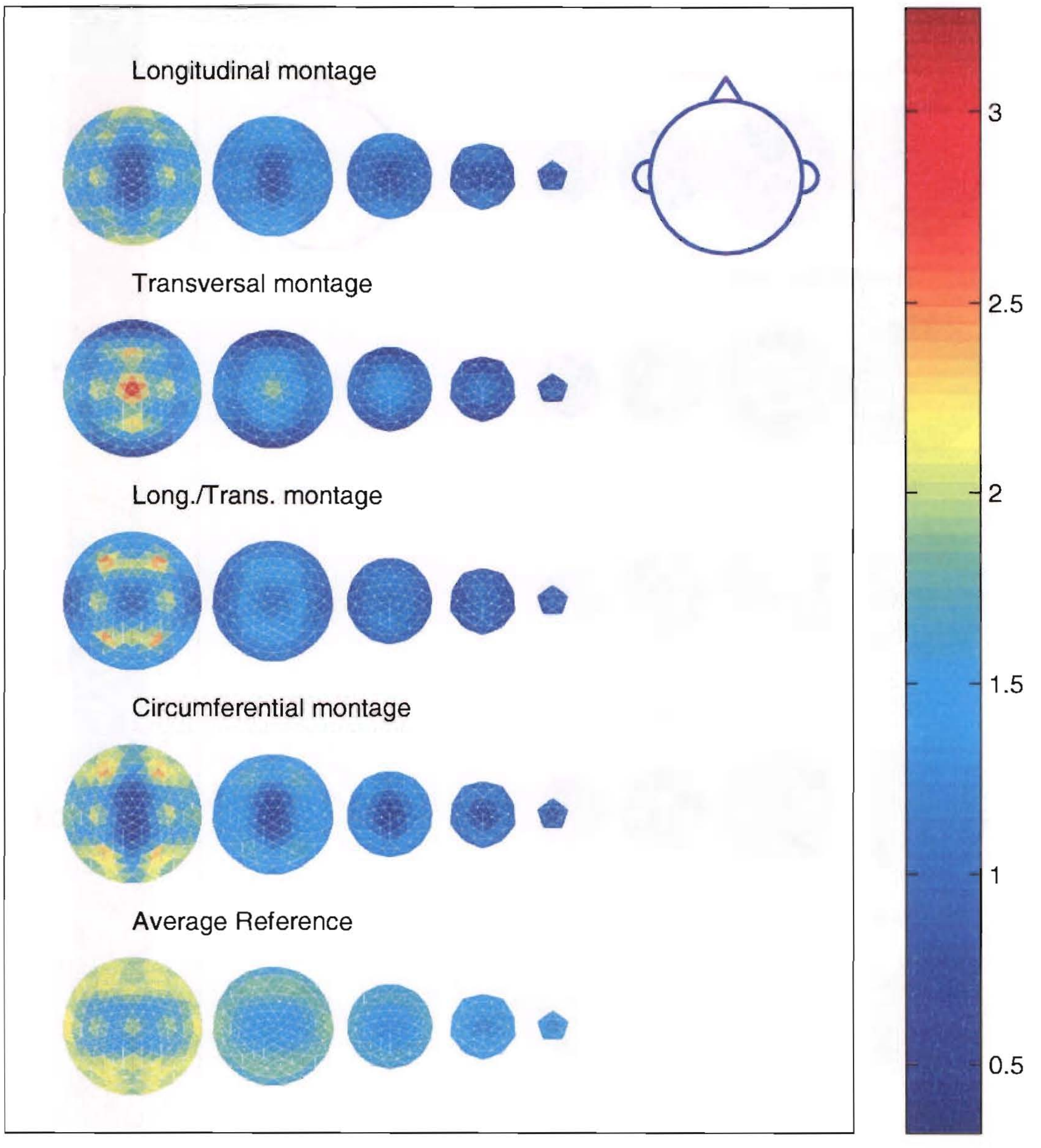

Figure 6.10 Sensitivity patterns $|\mathbf{v}_{\text{obs}}|$ in μV in response to the electric fields generated by radial dipoles of 10 nAm.

patterns for the longitudinal, transversal and longitudinal/transversal montages partly complement each other. The circumferential montage covers an area similar to the longitudinal one but shows higher sensitivity at the shell rims and a more pronounced gap in the centre. Mean sensitivities within the five shells are listed in Table 6.3.

Table 6.3 Mean $|\mathbf{v}_{\text{obs}}|$ in μV for *radial* sources on the five shells of the geodesic grid. Listed are the mean responses for each shell.

Montage	Eccentricity				
	7.0 cm	5.6 cm	4.2 cm	2.8 cm	1.4 cm
Longitudinal	1.4937	1.1819	1.0122	0.8658	0.8019
Transversal	1.4989	1.1465	0.9841	0.8324	0.7791
Long./Trans.	1.5054	1.1648	0.9959	0.8418	0.7829
Circumferential	1.6000	1.2803	1.1123	0.9483	0.8748
Average Reference	1.8593	1.6545	1.5398	1.4183	1.3629

Among the bipolar montages, the circumferential one is most sensitive to deep radial dipoles. However, for other dipole orientations it appears to be the least sensitive montage (Table 6.2).

6.7 LOCATION AND ORIENTATION ESTIMATES

In many source localization approaches the emphasis is put on confining the most probable area of activity. However, epileptiform discharges may occur at any location unless a focal type of epilepsy has already been diagnosed. One of the most distinctive spatial features of epileptiform discharges is their association with a pool of surface negativity. Since a pool of surface negativity is caused by radially oriented sources as shown in Fig. 6.3, this feature is related to the estimated source orientation. Therefore, in this work the estimation of source orientation is of equal importance to identifying the actual location. Spherical coordinates are used because a source pointing to the surface is easily identified. In contrast, Cartesian coordinates do not represent this feature explicitly.

In the following subsections, two approaches for source localization and orientation estimation based on the geodesic dipole model are presented. The first approach makes use of the best matching node only, while the second approach attempts to weight contributions from several nodes.

6.7.1 Closest Node

The geodesic dipole model presented in the previous section was used to estimate source locations and orientations. For each dipole $i = 1 \dots M$ in the model the forward and inverse matrices $\mathbf{K}_i$ and $\mathbf{K}_i^+$ were calculated. For a given measurement $|\mathbf{v}_{\text{obs}}|$, the dipole with the smallest regenerated PRD r_i was chosen from the set of 370, thus

$$n = \operatorname{argmin}_i\, r_i \; = \operatorname{argmin}_i \left(\frac{|\mathbf{v}_{\text{obs}} - \mathbf{K}_i\, \mathbf{j}_i|^2}{|\mathbf{v}_{\text{obs}}|^2} \right)^{\frac{1}{2}} \tag{6.11}$$

which, by virtue of the properties of the pseudoinverse, can be shown to be minimized for each node i by the dipole moment $\mathbf{j}_i = \mathbf{K}_n^+ \mathbf{v}_{\text{obs}}$. Having determined n, the location was given by the corresponding grid node $\hat{\mathbf{r}} = \mathbf{r}_n$.

6.7.2 Weighted Nodes

A source may be too distributed to match the dipole model and is usually contaminated by background activity and measurement noise. Therefore, a single dipole from the grid may only lead to a moderate match and useful information from other (mostly adjacent) dipoles is ignored. In order to incorporate this information, nodes can be assigned a weight w_i to estimate location

$$\hat{\mathbf{r}} = \sum_{i=1}^{M} w_i\; \mathbf{r}_i \tag{6.12}$$

and mean orientation

$$\bar{\mathbf{j}} = \sum_{i=1}^{M} w_i\; \mathbf{j}_i \,/\, |\mathbf{j}_i| \tag{6.13}$$

with the following choice of weights based on the PRD of each dipole in the model

$$w_i = w_0 * e^{-(r_i/\alpha)^2} \tag{6.14}$$

and normalization factor w_0 such that $\sum_{i=1}^{M} w_i = 1$. The parameter α controls the contribution of all nodes to weights: for small α only very few of the smallest r_i will be assigned a significant weight while for larger α a larger population of nodes will contribute to the estimate.

6.7.2.1 Optimal Choice of Parameter Alpha

The parameter α can be optimized for the SNR of a measurement. To determine the optimal α, the approach was tested in a simulation with 3000 dipoles of random orientation and random location within the upper hemisphere (max eccentricity 7.8 cm) with values for α ranging from 0.06 to 0.40. For each choice of α, 3000 random dipoles were generated and evaluated. For each dipole the forward solution was obtained, channel noise (SNR$=\infty$ dB or SNR$=$ 0 dB, respectively) was added and the weighted nodes approach was used to estimate the dipole's location and orientation. The cosine of the angle between true and estimated dipole orientation was used as a measure of accuracy. The longitudinal montage was applied because it is predominantly used in the recordings of the data base. The result is shown in Fig. 6.11. According to the

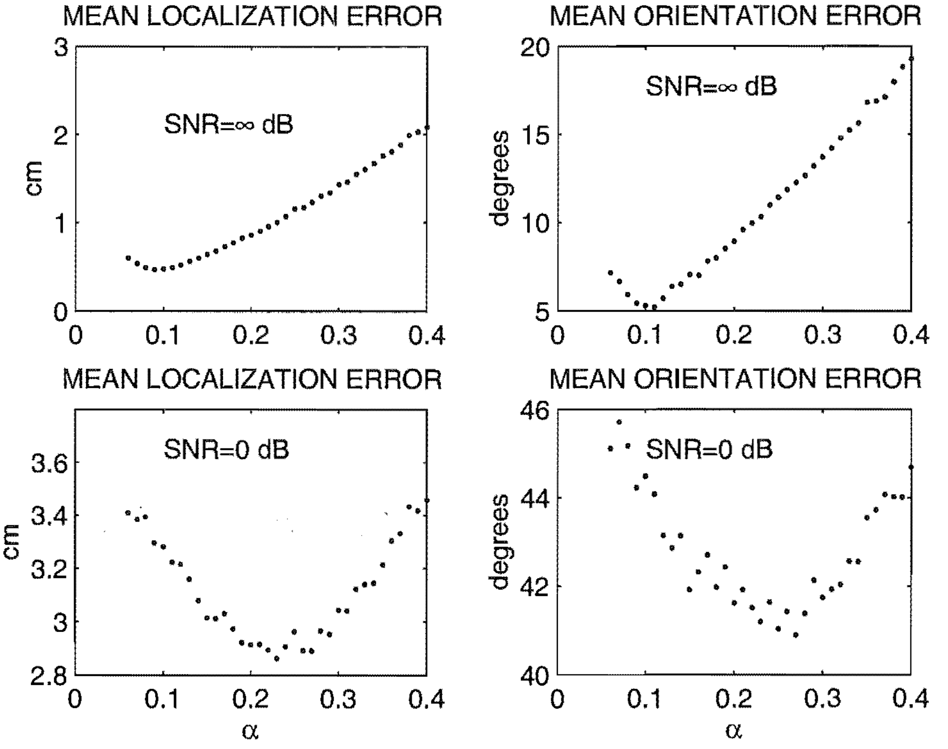

Figure 6.11 Mean localization error (left) and orientation errors (right) for the noiseless case (top row) and white Gaussian noise (bottom row, SNR= 0 dB).

simulation $\alpha = 0.09$ is optimal for localization in the noiseless case and $\alpha = 0.11$ leads to optimal orientation estimates. In the presence of noise (SNR= 0 dB) a choice of $\alpha = 0.27$ is optimal for localization and orientation estimates.

6.7.3 Performance

The performance of both approaches was tested for all montages with 1000 dipoles of random orientation and random location distributed as in the previous simulation. The errors of location and orientation estimates are shown in Fig. 6.12. The localization error peaks at about half the grid interval and shows a tail to the right which relates to locations that are not well covered by a specific montage. To simulate the effect of measurement noise, Gaussian white noise (SNR= 0 dB) was added to the channel vectors. The mean localization and orientation errors are listed in Table 6.4. On average, the referential montage features the smallest localization error and the best orientation estimate. The differences between montages range up to 54% for the localization error and 82% for orientation error. The results underline the advantages of a referential montage for source localization.

Table 6.4 Mean localization and orientation errors. Localization errors are given in mm, orientation errors in degrees, respectively. For the weighted nodes approach $\alpha = 0.09$ was used in the noiseless case and $\alpha = 0.27$ in the noisy case.

Montage	$\overline{\lvert\mathbf{r}-\hat{\mathbf{r}}\rvert}$				$\overline{\theta}$			
approach	closest		weighted		closest		weighted	
SNR (dB)	∞	0	∞	0	∞	0	∞	0
Longitudinal	8.9	37.7	4.9	29.5	10.4	47.4	5.7	42.7
Transversal	8.4	37.2	4.5	31.3	9.4	46.0	5.1	41.3
Long./Trans.	8.1	35.3	4.4	29.1	9.3	43.3	5.2	40.0
Circumferential	9.3	37.4	5.4	29.1	12.4	52.8	7.1	46.2
Average Reference	7.7	32.2	3.5	29.6	8.4	34.3	3.9	32.3

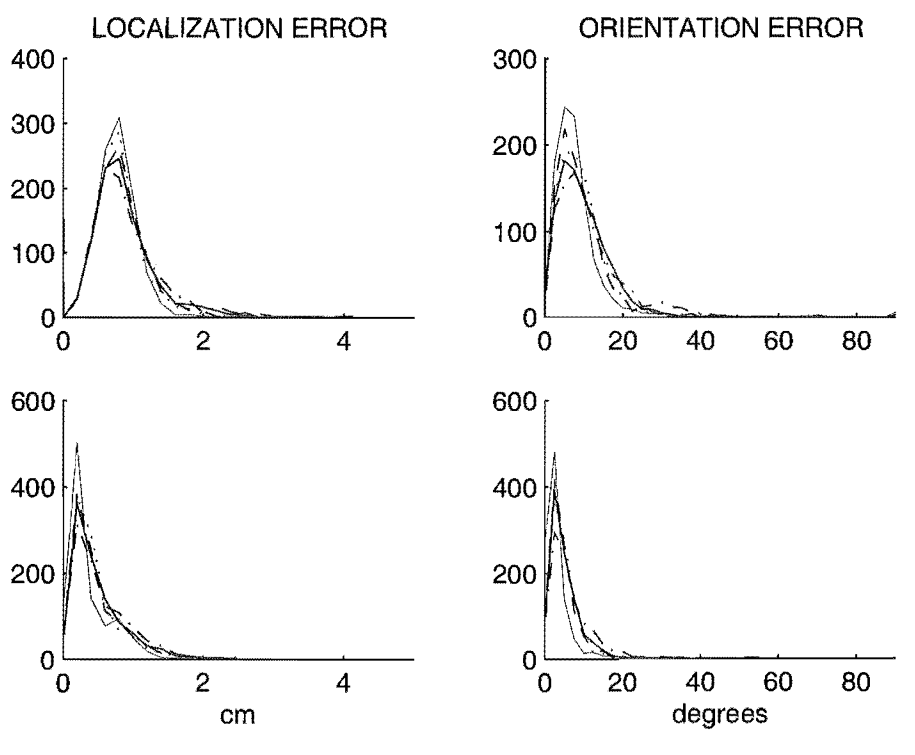

Figure 6.12 Distribution of location and orientation errors for five montages (in cm and degrees, respectively): longitudinal (solid line), transversal (dashed line), long./trans. (dotted line), circumferential (dash-dotted line) and average reference (thin line). Upper row shows results obtained for the closest node approach, lower row shows results obtained for the weighted nodes approach.

6.8 APPLICATION IN SPIKE DETECTION SYSTEM

The new approach to spatial analysis of the EEG is not meant to provide the best source location estimate for the noiseless case. Subsequent iterative optimization with, for example, the simplex method [Hara *et al.* 1999] would improve the performance of the procedure. However, for application in a spike detection system, it is important to obtain a fast and robust estimate for noisy measurements applicable to routine real-time screening of possible EDs.

It should be noted that α can be adjusted to optimize features other than source localization or orientation. Specifically, it can be adjusted to optimize the discrimination (ANOVA F-ratio) of epileptiform discharges from background activity. However, this application-specific choice for α requires the framework of the spike detection system and the training data set and therefore is presented in the following chapter.

6.9 SUMMARY

Essential concepts in the spatial analysis of the EEG were introduced. Standard methods to solve the electrostatic forward problem were presented.

One method was applied to illustrate the relationship between source space and channel space in several ways. The surface potentials generated by a radial and tangential current sources were illustrated. The concept of the lead field was introduced and related to the sensitivity of a single channel of the EEG. Subsequently, the spatial sensitivity of multichannel EEG was analyzed for several montages.

A number of current approaches to the inverse problem were discussed. A new application-specific method for the estimation of source localization and orientation based on a geodesic grid of dipoles was presented and evaluated in both noiseless and noisy environments.

Chapter 7

A WAVELET-BASED SPIKE DETECTION SYSTEM

7.1 INTRODUCTION

A wavelet-based two-stage detection system (henceforth referred to as "the system") for epileptogenic spikes is proposed. Initially, a statistical model for the background distribution of the wavelet coefficients is presented. From this statistical model the detection threshold is derived. Stage 1 of the system detects transients with a single-scale CWT using the detection threshold in a time-adaptive fashion.

In Stage 2, transients are analyzed by a blackboard approach: in a 300 ms window around the detection a more comprehensive 11-scale CWT of all channels is calculated. Discriminative single-channel features are extracted from the ridges and multiscale edges of the CWTs. Several multichannel features are derived from the spatial distribution of the single-channel features. Fisher's linear discriminant with the Bayesian approach, as outlined in section 5.5, is applied to the single-channel features to obtain the probability that each waveform resembles a spike. Additional multichannel features are derived from these probabilities. Another linear discriminant function is subsequently applied to the multichannel features to obtain a final classification.

The training data set used in the design of the system is presented with the test data set and the final results in the next chapter.

7.2 WAVELET FILTERS

The spike detection system makes use of the CWT with eleven even integer scales as listed in Table 7.1. The psi_1 wavelet is used as described in section 3.4.4. The series

Table 7.1 Scales and corresponding centre frequencies f_k used in the CWT. Scales are listed as filter lengths in samples. Centre frequencies are given for the psi_1 wavelet used. Scale 18 is used for the initial detection in the first stage.

index k	1	2	3	4	5	6	7	8	9	10	11
scale a_k	6	8	12	18	24	36	48	72	96	144	192
f_k (Hz)	50.0	37.5	25.0	16.7	12.5	8.3	6.25	4.167	3.1250	2.0833	1.5625

of scales approximates a harmonic progression with two scales per octave, rounded to even integers. True amplitude normalization was used as define in sections 3.5.1 and 4.3.1. The bandwidth covered by each scale is illustrated in Fig. 7.1. All relevant parts

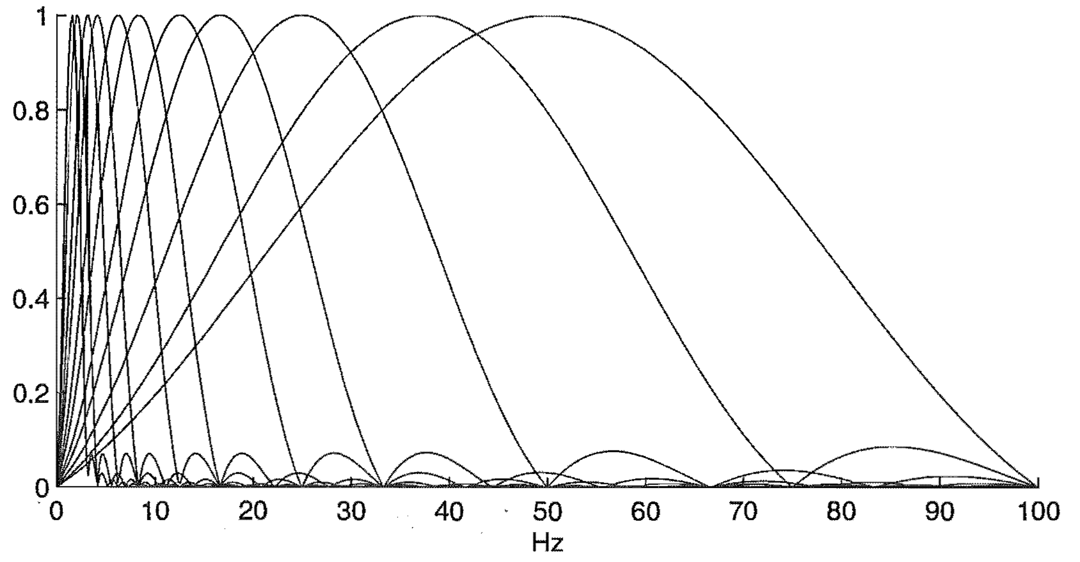

Figure 7.1 Spectral coverage of the CWT with scales used with the psi_1 wavelet. True amplitude normalization is used.

of the spectrum are covered with some redundancy. Scale 24 features the same centre frequency as the alpha band (12.5 Hz) but covers a wider bandwidth. Scale 18 is used for the initial detection in Stage 1 whereas the complete set of scales is analyzed in Stage 2 of the system.

Scale 18 has been chosen for the initial detection in Stage 1 because a minimal number of false detections was achieved while all EDs were detected (the selection criterion was based on the confusion matrix, Eq. 5.9). However, the choice of scale is not critical, as similar results were obtained for scale 12 and 24 in Stage 1. A criterion based on the F-ratio could also be employed in the optimization of Stage 1.

The range of scales for the Stage 2 analysis was chosen to cover the entire frequency range of the EEG of clinical significance and to provide sufficient frequency resolution to track ridges in the CWT. With only one scale per octave the frequency resolution of the CWT would be too coarse to represent the chirp features (see below), while more than two scales per octave would represent unnecessary redundancy for the CWT.

7.2.1 Background Distribution

A single scale of the CWT relates to a limited band within the EEG signal. Within the context of the limited bandwidth, the statistical behaviour of the background EEG can be approximated with a given distribution. For example, the distribution of wavelet coefficients of a 214-s epoch of background activity from an occipital bipolar channel

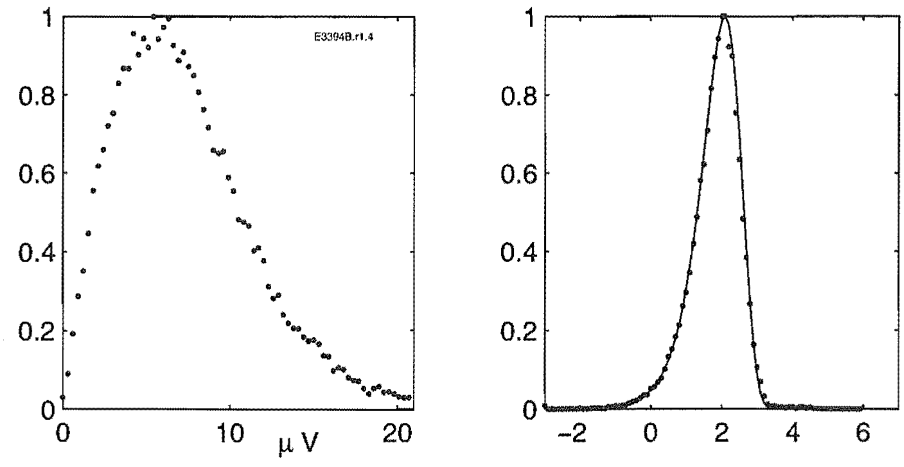

Figure 7.2 Amplitude distribution of wavelet coefficients resulting from a single channel of background EEG and the Psi-1 wavelet. Distribution of modulus (left) and log modulus (right) with fitted Chi-Square distribution (df=3.5). 42808 samples (214 s) were evaluated.

is shown in Fig. 7.2 (left). A chi-square distribution can be fitted to the magnitude distribution of the wavelet coefficients. The *degrees of freedom* N of the chi-square distributed variable x can estimated by

$$N = \frac{2\mathbf{E}[x]^2}{\mathbf{E}[x^2] - \mathbf{E}[x]^2} \tag{7.1}$$

giving values in the range of $N = 2.0 \ldots 6.8$ [Senhadji *et al.* 1994].

Taking the log magnitude of wavelet coefficients leads to a distribution which is less skewed (Fig. 7.2 right). In this representation, N can be adjusted to match the mean, standard deviation and skewness to the log distribution. The best match is found for $N = 3.5$. The quality of this approximation can be observed to be remarkable. The distribution of wavelet coefficients of artifact-free background activity is remarkably consistent across scales (Fig. 7.3). This consistency has also been shown to hold across channels and individuals. Thus, the log wavelet coefficient of a sharp transient can be related to the mean of the log background distribution and, hence, can be used to estimate the probability that a given wavelet coefficient occurs by chance.

The mean of the log distribution relates to the mean amplitude in each frequency band. In contrast, the standard deviation and skewness of the log distribution are independent from the amplification factor and relate to the nonlinear dynamics of the EEG. The CWT is applied in several ways in the system:

- it represents a filter for transient detection which adapts to the changing background EEG (5-s mean of log magnitude),
- it acts as a multiscale edge representation to establish the sharpness of transients, and

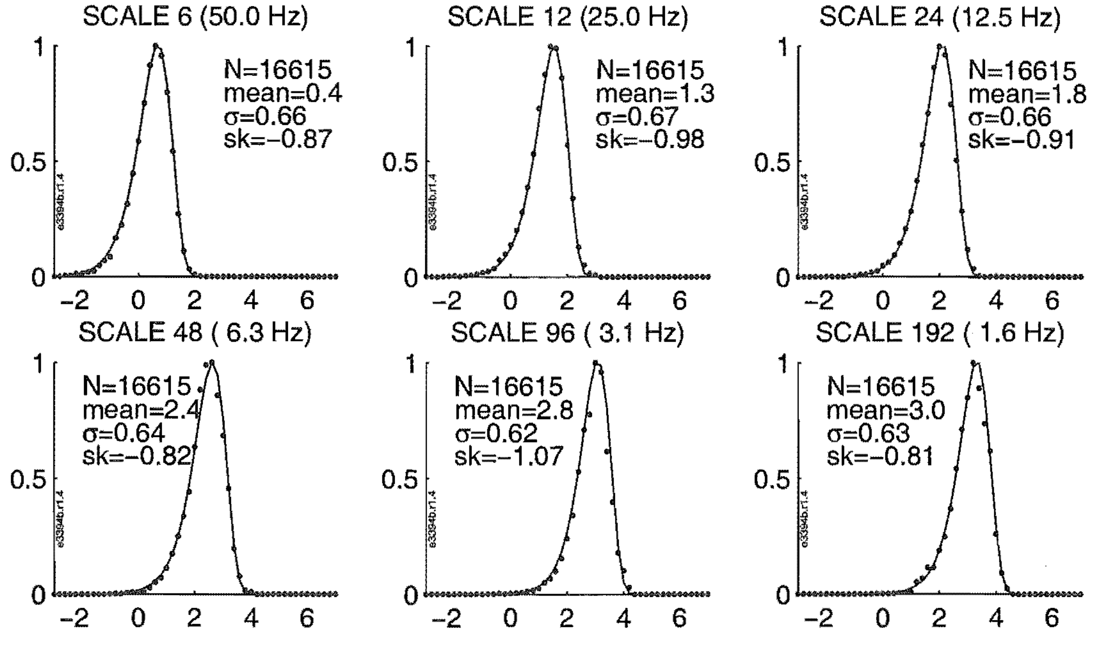

Figure 7.3 Distributions of log magnitude wavelet coefficients for 6 scales (1.6 Hz to 50 Hz). The mean, standard deviation and skewness of each distribution is given. 16615 samples corresponding to 83 s of artifact free background EEG were evaluated.

- it is used as a time-frequency representation in the estimation of the instantaneous frequency of oscillations.

7.3 SYSTEM SCHEMATIC

The system consists of two stages as shown in Fig. 7.4. Stage 1 is designed to detect multichannel transients while rejecting EMG and EOG artifacts. The single-channel analysis is detailed in Fig. 7.5, while the Stage 1 multichannel detector is illustrated in Fig. 7.6. In Stage 2 the multichannel-multiscale context of transients or *candidate epileptiform discharges* (CEDs) is analyzed in detail and classified as ED or artifact with discriminant functions. Details of the Stage 2 single-channel feature extraction are illustrated in Fig. 7.8, while the complete Stage 2 multichannel analysis is detailed in Fig. 7.9.

7.4 STAGE 1: TRANSIENT DETECTION

A multichannel transient or CED is detected as a variation in the magnitude of the wavelet coefficients of a specific scale. If two or more channels show a significant variation and no muscle or eye activity is present, a CED is deemed to have been detected.

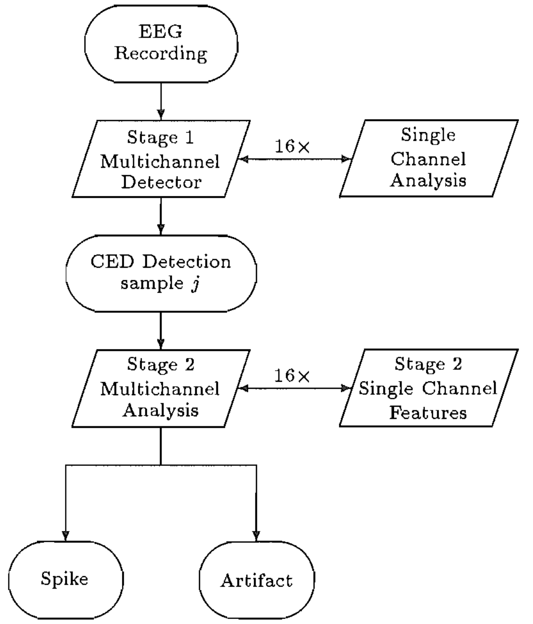

Figure 7.4 The system consists of two stages: Stage 1 detects multichannel transients and rejects EMG and EOG activity. The multichannel-multiscale context of detections is scrutinized in Stage 2 where the final classification as epileptiform spike or artifact is made.

7.4.1 Single-Channel Processing

In the following, the EEG is represented by a set of sampled potentials v_{ij} on channel i and sample index j. Initially, each channel is filtered with a complex wavelet filter (psi_1 wavelet, scale 18, centre frequency 16.7 Hz) as illustrated in the left branch of Fig. 7.5. The magnitude of wavelet coefficients is used since the phase of the transients is unknown. In the next step, the log of the magnitude coefficients is taken to reduce the skewness of the distribution as described in section 7.2.1. Since the log coefficients' distribution is closer to a Gaussian than that of the raw coefficients, they provide a more stable estimate of the mean background activity [Gölz *et al.* 1999]. The mean of the log coefficients of the 5 s (1000 samples) preceding the current sample is subtracted to adapt to changes in the background activity. Such a 5-s interval was suggested by Gotman [1985] to define the context for automatic spike detection.

In parallel to the wavelet filter, a high-pass filter is used to detect artifacts caused by the EMG within a single channel (right branch in Fig. 7.5). The high-pass filter is of four-tap FIR form specified by

$$P_{[\mathrm{HP}]}(z) = z^{-2} + 3z^{-1} + 3 + z \; . \tag{7.2}$$

EMG activity is measured by the 100-ms average (floating mean of 20 samples centered around the current sample) of the magnitude of filtered samples. If a threshold of

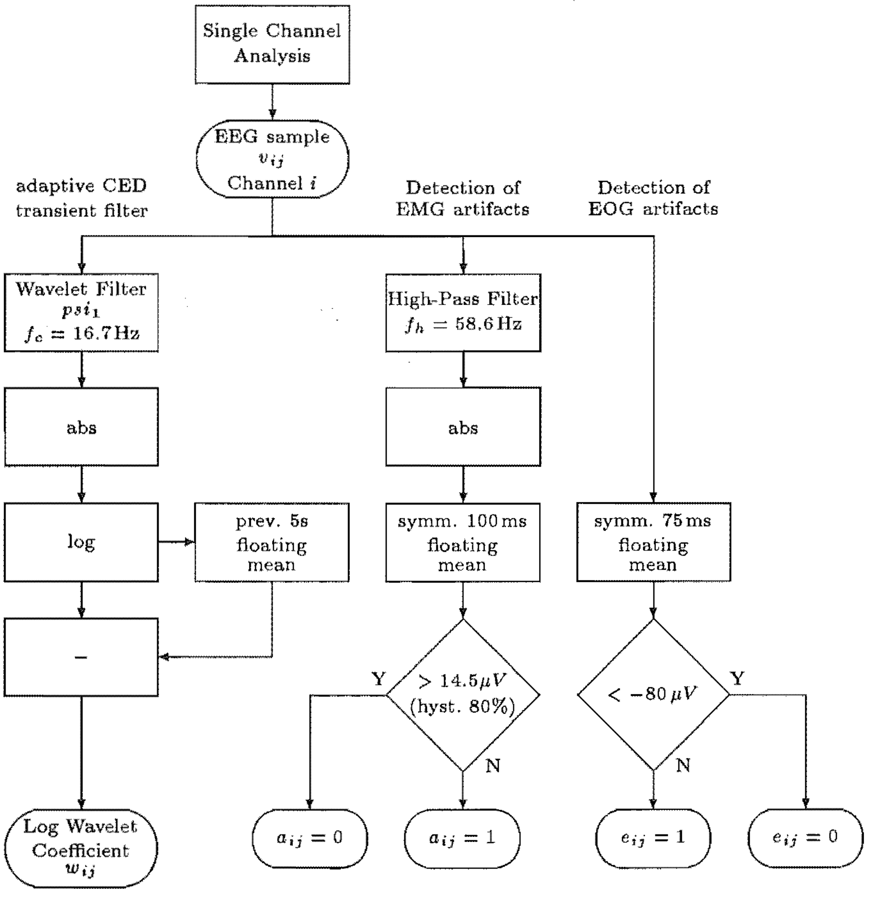

Figure 7.5 Single-channel analysis: wavelet filter with adaptive background processing (left branch) EMG artifact detection (centre branch) and EOG artifact detection (right branch).

14.5 μV is exceeded, the current sample is marked as an EMG artifact ($a_{ij} = 0$). A value of $a_{ij} = 1$ indicates that channel i is, at least, relatively free from EMG at sample j. The threshold is chosen to allow for small amplitude EMG activity to be accepted. However, strong EMG activity that could trigger the detection of a CED is rejected. Once the EMG detection threshold is exceeded, it has to drop below 80% of the threshold for the output to return to the non-EMG state (hysteresis characteristic).

Artifacts caused by the EOG are detected if the potential drops below a threshold of -80 μV with respect to the 75-ms average in several eye channels [Dingle 1992]. A value of $e_{ij} = 1$ indicates that channel i is relatively free from EOG activity at sample j.

7.4.2 Multichannel Transient Detection

Following the single-channel processing and possible rejection of samples due to artifacts, the wavelet coefficients of all accepted channels are considered in the transient detector (Fig. 7.6). Since a CED can only be confirmed to be a ED if it is observed in two or more channels, the largest and the *second largest* wavelet coefficient of the accepted channels are determined for each sample j

$$x_j = \max_i w_{ij}\ a_{ij} \tag{7.3}$$

and

$$y_j = \max_{i\,:\,w_{ij} < x_j} w_{ij}\ a_{ij}\ . \tag{7.4}$$

Note that the EMG indicators a_{ij} force EMG contaminated channels to zero. Thus, the two channels with the strongest signal amplitude within the wavelet's pass-band are accumulated in x_j and y_j. While the distributions of w_{ij} of a specific channel i show a negative skewness, as seen in Fig. 7.3, the x_j and y_j feature a positive skewness (Fig. 7.7) since they accumulate the transient features of all channels. To accommodate changes in the background activity, the mean of the preceding 5 s is subtracted from the x_j and y_j. Subsequently, separate detection thresholds are applied to x_j and y_j.

While EMG artifacts can be identified in each individual channel, EOG artifacts need confirmation from several frontal channels. The number of channels contaminated by EOG is derived from the sum of the inverted EOG indicators

$$z_j = \sum_i (1 - e_{ij})\ \hat{e}_i \tag{7.5}$$

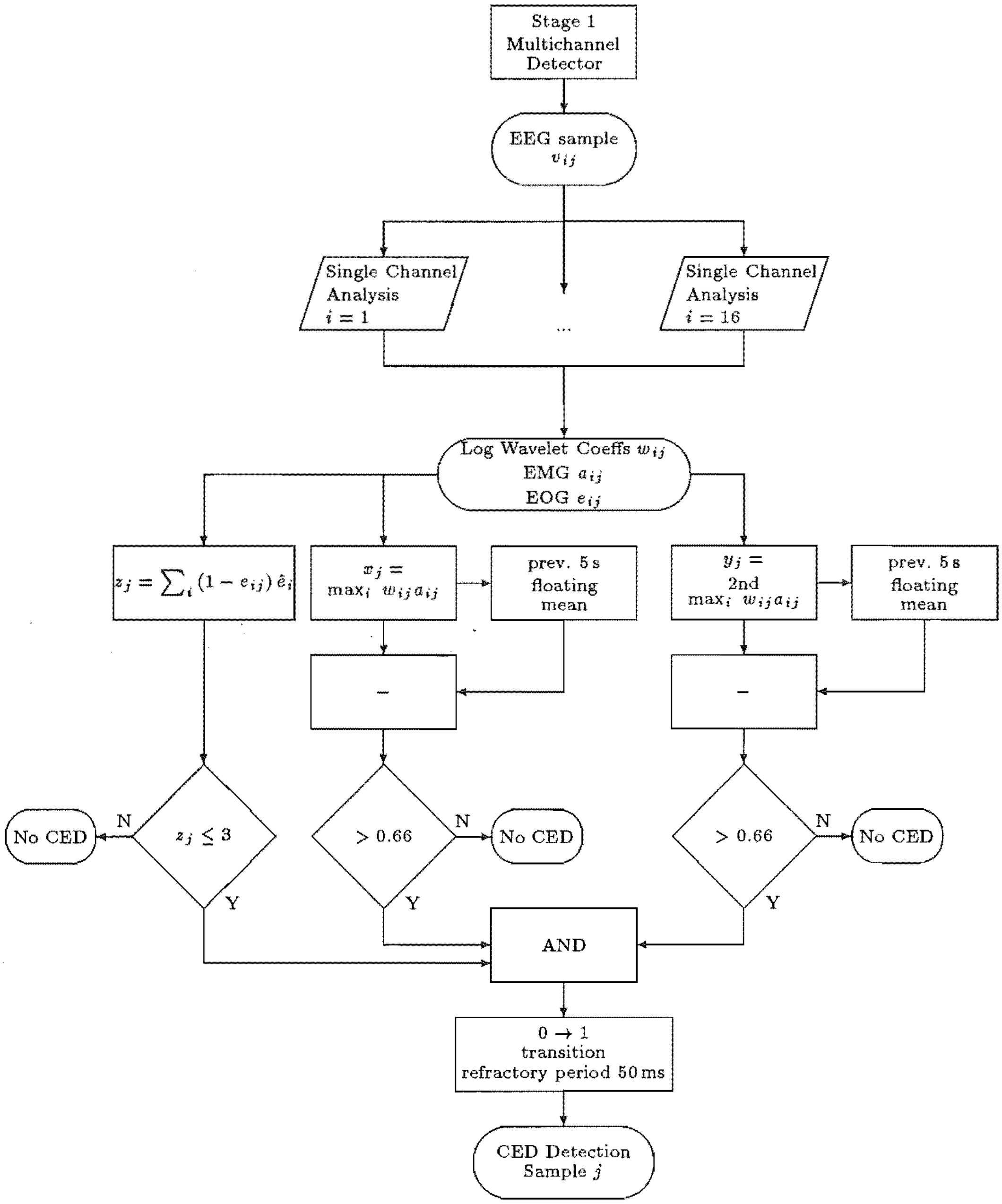

Figure 7.6 Multichannel transient detector.

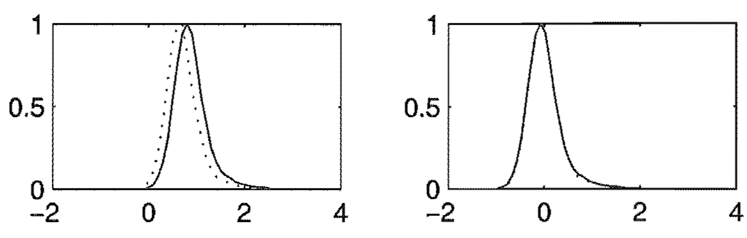

Figure 7.7 Left: distribution of x_j (solid line) and y_j (dotted line). Right: subtraction of the 5-s floating mean shifts both peaks to the centre.

where

$$\hat{e}_i = \begin{cases} 1 & \text{if channel } i \text{ is a frontal channel} \\ 0 & \text{otherwise.} \end{cases} \tag{7.6}$$

An EOG artifact (eye-blink or eye movement) is detected if three or more frontal channels contain EOG activity ($z_j \geq 3$).

If detection thresholds are exceeded for x_j and y_j at sample j, but are not exceeded for the previous sample $j-1$, a CED is deemed to have been detected. Once a CED detection has occurred, no further detections are made within a 50 ms (10 sample) refractory period.

7.5 STAGE 2: TRANSIENT ANALYSIS AND CLASSIFICATION

Each CED detected by Stage 1 is scrutinized more fully in Stage 2. All available information relating to the CED is gathered on a 'blackboard' that covers time, scale and space.

7.5.1 Synopsis

The full eleven-scale CWT is obtained for all channels (refer to Table 7.1) and stored in a 3D array of wavelet coefficients called a *blackboard*. The fiducial point of the multichannel transient is derived from the wavelet coefficients within the blackboard. Several features are extracted from the blackboard which help discriminate epileptiform transients from alpha spindles and beta bursts. The processing steps of the single-channel feature analysis of Stage 2 is illustrated in Fig. 7.8. The Euclidian distance between adjacent channels is evaluated to detect electrode artifacts. Single-channel features are extracted and fed into a linear discriminant function. The output of the discriminant function is evaluated with a Bayesian approach, resulting in a set of single-channel spike probabilities.

Four multichannel features are derived from the spatial potential distribution of the CED and the single-channel spike probabilities (Fig. 7.9). The spatial potential distribution is analyzed with respect to a dipole model and with a spatial filter which is sensitive to the adjacency patterns of focal EDs. Multichannel features are eventually fed into a second linear discriminant function, the output of which is again interpreted with a Bayesian approach. The final CED classification is based on the multichannel spike probability.

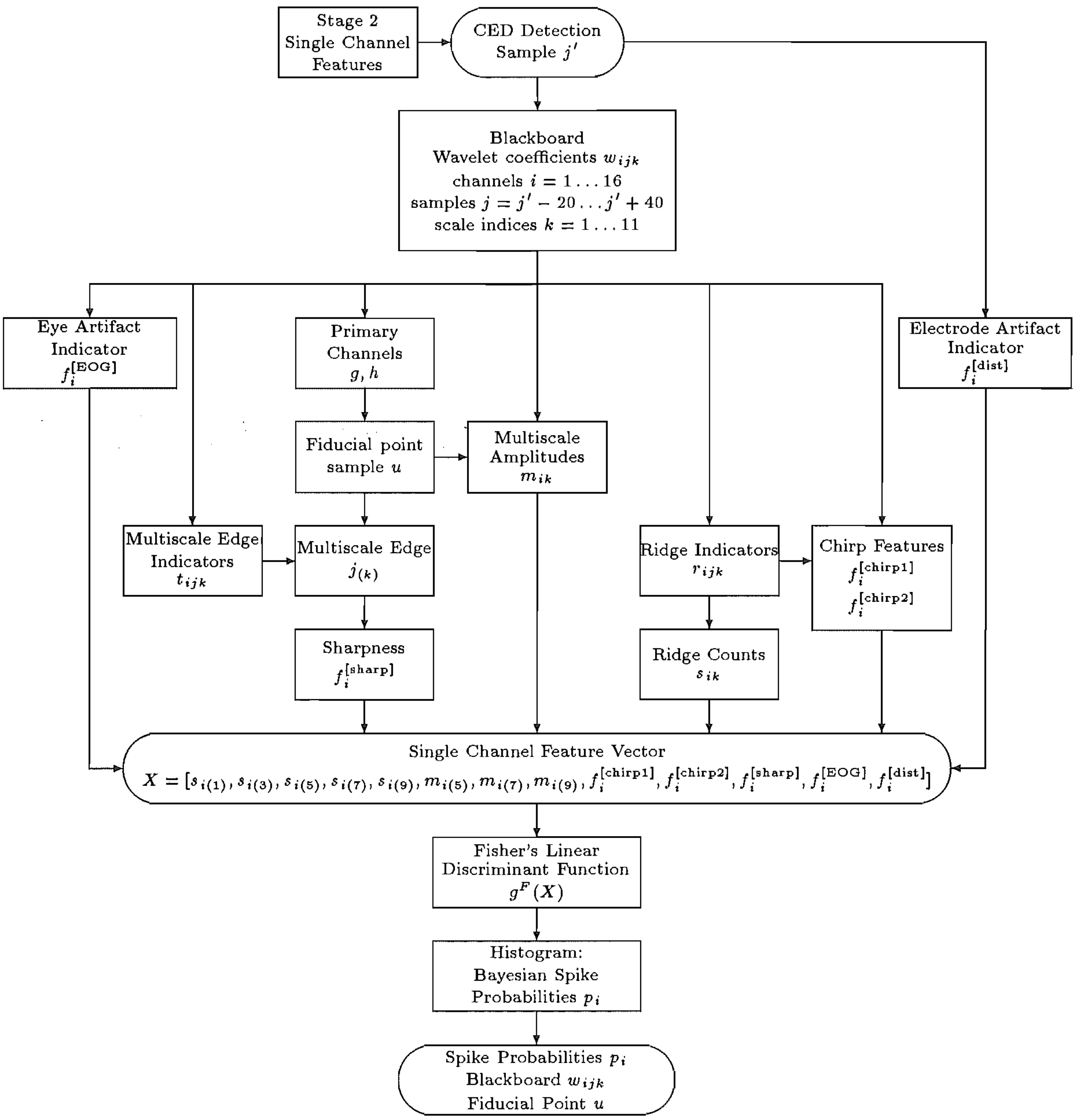

Figure 7.8 Extraction and evaluation of single-channel features of a transient from the wavelet coefficients of all channels and scales ('blackboard'). An additional electrode artifact indicator is derived directly from the EEG signals.

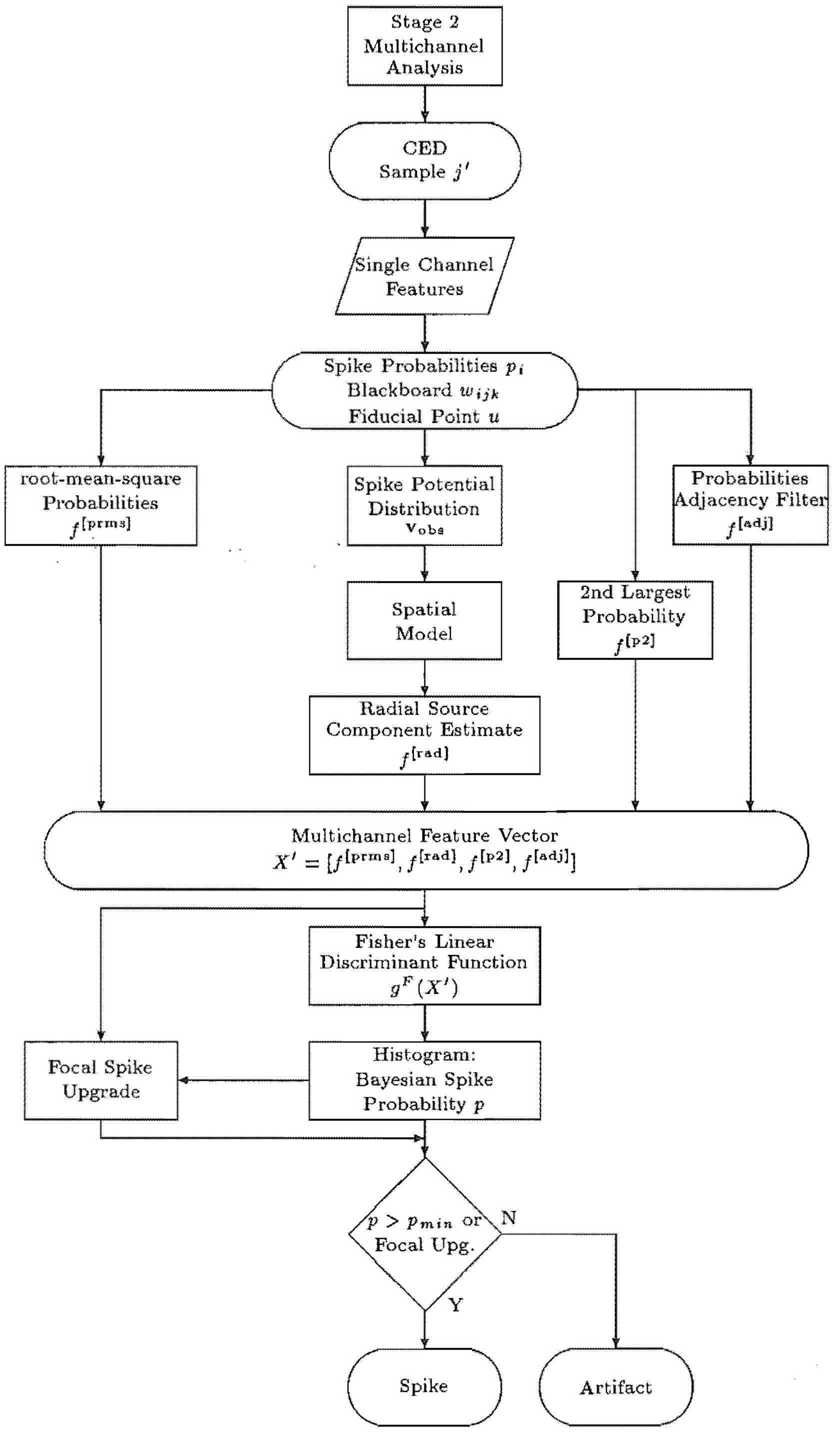

Figure 7.9 Extraction and classification of multichannel features of a CED.

7.5.2 Single-Channel Features

7.5.2.1 The Multichannel Multiscale Blackboard

For a 300-ms period (60 samples, 20 samples before to 40 samples after the Stage 1 detection point), the complete CWT covering eleven scales is calculated for all channels. Thus, the entire multichannel transient is put under the 'mathematical microscope'. For channel i, sample j and scale index k the complex wavelet coefficients w_{ijk} are

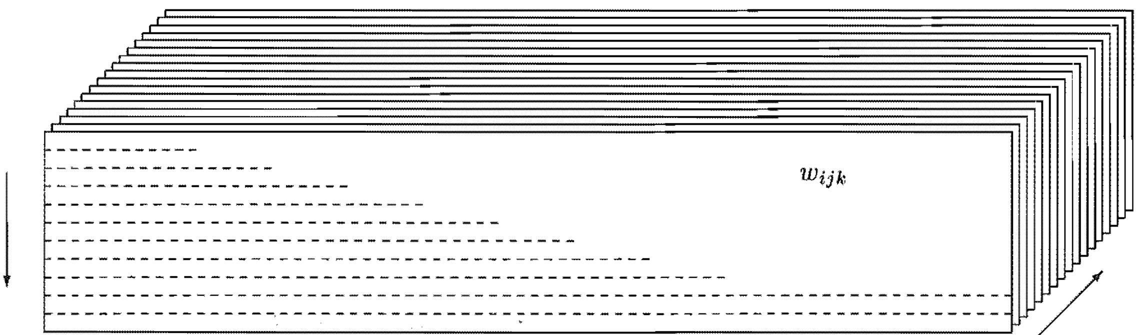

Figure 7.10 Blackboard: 3D array of complex wavelet coefficients w_{ijk} for channel i, sample j and scale index k. The Stage 1 detection is aligned to sample $j = 20$.

obtained from a discrete convolution with the amplitude-normalized psi_1 wavelet filter of length a_k (Table 7.1, Eq. 3.10). This 3D array of wavelet coefficients is the basis for the subsequent analysis and is referred to as the blackboard. It should be noted that, because of the width of the wavelet at the maximum scale, the wavelet coefficients within the blackboard relate to EEG samples that occur up to 96 samples earlier or later than the blackboard. The blackboard covers 16 channels, 60 samples and 11 scales, amounting to 10560 complex wavelet coefficients. The Stage 1 detection point corresponds to sample $j = 20$ of the blackboard. All features that are used to identify a spike are derived from the blackboard. Additional features are used to reject electrode artifacts on adjacent channels. Please note that in the previous section (description of the Stage 1 detector), w_{ij} represented the log wavelet coefficients - in this section they represent the complex wavelet coefficients.

7.5.2.2 Identification of Primary Channels and Fiducial Point

Two primary channels g and h for the *candidate epileptiform transient* (CET) are identified. Only the primary channels are subsequently used to locate the fiducial point u of the CET.

To find the primary channels, the mean square amplitude of 22 consecutive scale wavelet coefficients $w_{ij(3)}$ of the spike is determined for each channel in a window around the Stage 1 detection point

$$p_i = \frac{1}{22} \sum_{j=14}^{35} |w_{ij(3)}|^2 \ . \tag{7.7}$$

The two primary channels g and h of the spike are determined as

$$g = \text{argmax}_i \ p_i \tag{7.8}$$

and

$$h = \text{argmax}_{i \neq g} \ p_i \ . \tag{7.9}$$

The sample index, u, of the fiducial point of the CET is derived from the scale-12 wavelet coefficients of channels g and h using

$$u = \text{argmax}_j \ |\Re\{w_{gj(3)}\}|^2 + |\Re\{w_{hj(3)}\}|^2 \ . \tag{7.10}$$

7.5.2.3 Indicators derived from Multiscale Ridges

Dominant frequencies in the EEG can be tracked by identifying ridges in the wavelet transform. In the following, wavelet coefficients w_{ijk} are considered part of a ridge if the ridge indicator

$$r_{ijk} = \begin{cases} 1 & \text{if } |w_{ijk}| > |w_{ij(k-1)}| \text{ and } |w_{ijk}| > |w_{ij(k+1)}| \text{ and } |w_{ijk}| > w_0 \\ 0 & \text{otherwise.} \end{cases} \tag{7.11}$$

is non-zero or, in other words, there is a local maximum with respect to scale. The constant $w_0 = 8.0\,\mu\text{V}$ defines the threshold for significant EEG activity. Coefficients w_{ijk} of scales outside the defined range of $k = 1 \ldots 11$ are assumed to equal zero. The occurrence of non-zero ridges indicators for examples of epileptiform spikes, eye blinks and oscillatory activity is illustrated in Fig. 7.11. A characteristic feature of epileptiform spikes and spike-and-wave complexes is a series of about 12-20 non-zero ridge indicators in scales 12 to 24 near the fiducial point followed by adjacent non-zero ridge indicators in larger scales. In most cases of eye-blinks the $rij(11)$ (scale-192 ridge indicators) are triggered. Alpha spindles can occupy the same area as spikes but show a more symmetric pattern. The frequency of the alpha rhythm is not constant and can be described as a waxing and waning pattern. The corresponding ridge may alternate between scales 24 and 36 with 8.3 and 12.5 Hz centre frequencies respectively. A beta burst rarely triggers a Stage 1 detection and is usually confined to scales 12 and 18. Muscle activity can spread from scale 6 down to scale 18 and features very incoherent oscillations.

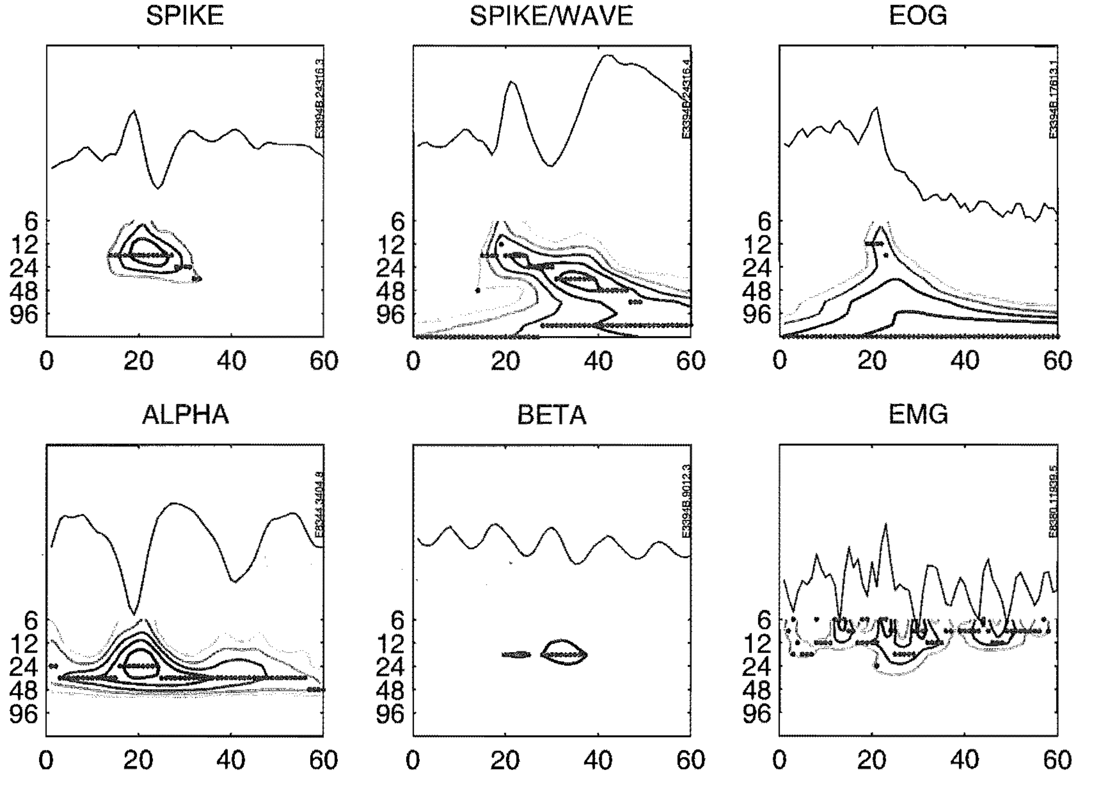

Figure 7.11 Behaviour of ridge indicators in various situations: epileptiform spike (top left), spike and wave complex (top centre), eye-blink with sharp onset (top right), alpha spindle (bottom left), beta burst (bottom centre) and muscle activity (bottom right). 60 samples of the signal are plotted above the wavelet coefficients. Contour lines ($8\,\mu$V spacing) show $|w_{ijk}|$, non-zero ridge indicators r_{ijk} are marked with dots. Note that all wavelet coefficients above scale 48 are influenced by samples outside the plotted interval.

Table 7.2 Ridge counts s_{ik} used in the system and corresponding frequency ranges. Also listed are oscillations and transients that cause substantial ridge counts.

indicator	EEG band	transients	scale	frequency range
$s_{i(1)}$	gamma	EMG	6-8	31.3 Hz and above
$s_{i(3)}$	beta	spike	12-18	14.6-31.3 Hz
$s_{i(5)}$	alpha	spike/wave	24-36	7.3-14.6 Hz
$s_{i(7)}$	theta	wave	48-72	3.7-7.3 Hz
$s_{i(9)}$	delta	EOG	96-144	1.8-3.7 Hz

For each channel i and pair of adjacent scales k and $k+1$, the samples positioned on ridges are counted to form the ridge count

$$s_{ik} = \sum_{j=1}^{N_s} \sum_{k'=k}^{k+1} r_{ijk'} \tag{7.12}$$

where $N_s = 60$ is the number of samples covered by the blackboard. Since the ridge indicator r_{ijk} can only be non-zero on either scale k or scale $k+1$, $0 \leq s_{ik} \leq 60$. For example, $s_{i(5)}$ can be considered an indicator for alpha activity. Even moderate alpha activity exceeds the threshold w_0 and leads to a significant count of ridge samples. All ridge counts s_{ik} used in the system and the corresponding EEG frequency bands are listed in table 7.2.

7.5.2.4 Multiscale Edges

Ridge indicators capture the centre frequency of oscillatory and transient activity in the scalogram only. However, other parts of the time-scale distribution of a CED near the fiducial point also feature significant amplitudes. Features not captured by ridges are captured by *multiscale edges.*

The multiscale edge indicator t_{ijk} defines temporal local maxima of the wavelet coefficients, thus

$$t_{ijk} = \begin{cases} 1 & \text{if } |w_{ijk}| > |w_{i(j-1)k}| \text{ and } |w_{ijk}| > |w_{i(j+1)k}| \\ 0 & \text{otherwise.} \end{cases} \tag{7.13}$$

The multiscale edge is traced from the fiducial point on scale 12 to scale 6. The sample j_3 of the multiscale edge at scale 12 corresponds to the largest temporal local maximum near the fiducial point u, thus

$$j_3 = \operatorname{argmax}_{j=u-5}^{u+5} t_{ij(3)} \left|w_{ij(3)}\right| . \tag{7.14}$$

The multiscale edge was traced recursively from scale 12 to scale 6 by

$$j_k = \operatorname{argmax}_{j=j_{(k+1)}-5}^{j_{(k+1)}+5} t_{ijk} \left|w_{ijk}\right| , \; k = 2, 1. \tag{7.15}$$

The behaviour of multiscale edge indicators and the traced multiscale edge is illustrated in Fig. 7.12. The features of the scalogram highlighted by the multiscale edge complement those features covered by the ridge indicators. While ridge indicators are closely related to the instantaneous frequency of a transient, multiscale edges are more closely related to time-domain features such as slope. The multiscale edge can be used to mark the sharpest slope of the transient. In practice, only the amplitude of the scale-6 wavelet coefficient of the multiscale edge is used in the system as an indicator

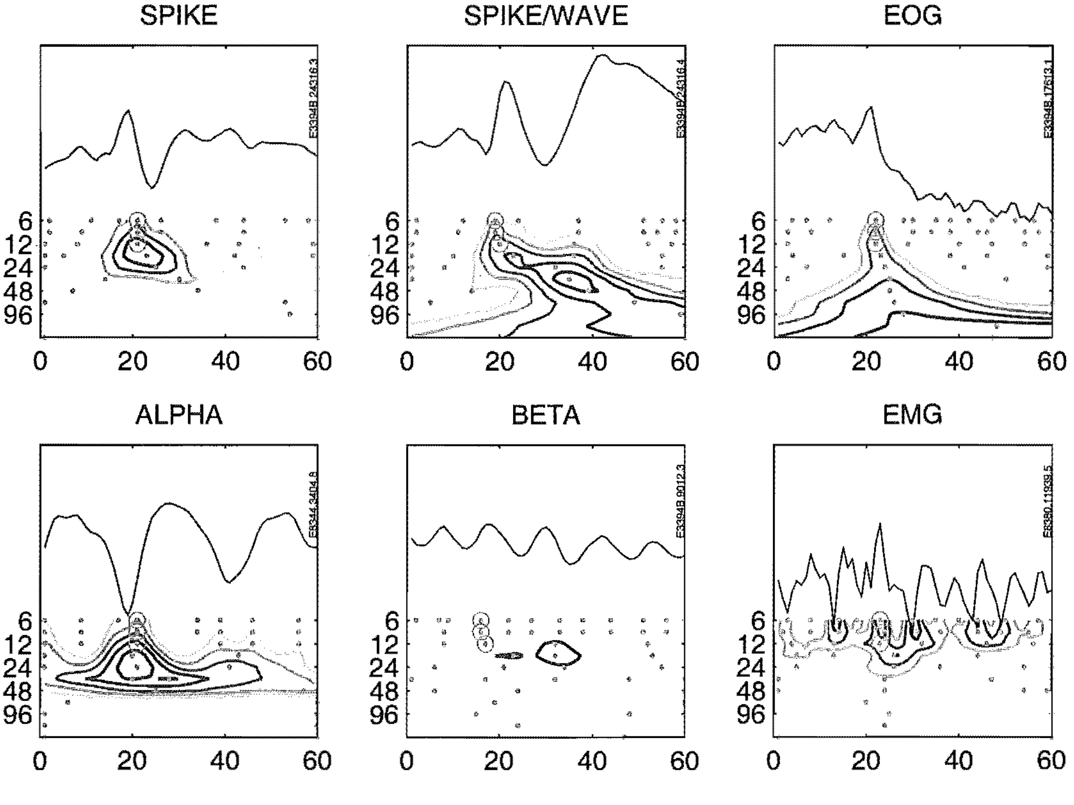

Figure 7.12 Behaviour of multiscale edge indicators in the same situations as in Fig. 7.11. Contour lines (8 μV spacing) show $|w_{ijk}|$, whereas non-zero multiscale edge indicators t_{ijk} are marked with dots. The multiscale edge closest to the fiducial point has been traced from scale 12 to scale 6 (marked by circles).

of sharpness of the transient, thus

$$f_i^{[\text{sharp}]} = |w_{i(j_1)(1)}| \ . \tag{7.16}$$

7.5.2.5 Multiscale Amplitude Features

For each channel i and pair of adjacent scales k and $k+1$, the mean amplitude around the fiducial point is determined by

$$m_{ik} = \frac{1}{2\sigma N_s} \sum_{k'=k}^{k+1} \sum_{j=u-10}^{u+10} |w_{ijk'}| \tag{7.17}$$

with the RMS amplitude of artifact free EEG given by

$$\sigma^2 = \frac{\sum a_{ij} \ e_{ij} \ v_{ij}^2}{\sum a_{ij} \ e_{ij}} \ . \tag{7.18}$$

The mean amplitudes $m_{i(1)}$, $m_{i(3)}$ and $m_{i(5)}$ covering three non-overlapping rectangular areas of the scalogram (scales 6 to 36) are used as features in the system.

7.5.2.6 EOG Indicator

The amplitude of the scale-192 wavelet coefficient 20 samples after the primary detection point captures slow variations of the potential such as EOG activity. It is therefore used as the EOG indicator

$$f_i^{[\text{EOG}]} = |w_{i(40)(11)}| \ . \tag{7.19}$$

7.5.2.7 Chirp Features

The CWT of epileptiform spikes and spike-and-wave complexes shows high frequency components followed by lower frequency components. Generally, interictal epileptiform activity shows time-scale features similar to a *chirp* signal with a small number of oscillations. Since the energy of such a chirp is concentrated on a diagonal or curved line in the scalogram, two 2D filters are proposed to capture this specific feature. Both filters integrate wavelet coefficients on a path at various translations. The scale indices c_j, $j = 1 \ldots 19$, of the path in the time-scale plane follow the sequence

$$c = \{2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5\} \tag{7.20}$$

as illustrated in Fig. 7.13. The first chirp filter can be defined as

$$f_i^{[\text{chirp1}]} = b_1 \ \max_d \sum_j \ |w_{ij(c_{(j-d)})}| \ . \tag{7.21}$$

with normalization factor

$$b_1 = 1/\sum_{j=1}^{60}\sum_{k=2}^{5} |w_{ijk}| \ . \tag{7.22}$$

The second chirp filter takes only wavelet coefficients on ridges into account:

$$f_i^{[\text{chirp2}]} = b_2 \ \max_d \sum_j r_{ijk}|w_{ij(c_{(j-d)})}| \tag{7.23}$$

with normalization factor

$$b_2 = 1/\sum_{j=1}^{60}\sum_{k=2}^{5} r_{ijk}|w_{ijk}| \ . \tag{7.24}$$

Both filters sweep the scalogram from left to right. At each translation, amplitudes

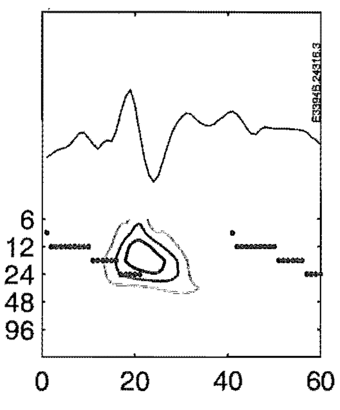

Figure 7.13 Dotted lines indicate the path of the chirp filters as defined by Eq. 7.20. The path is shown for leftmost and rightmost translation. The filter output corresponds to the maximum line integral along the path for all translations.

of wavelet coefficients are integrated along the path defined by the c_j. The maximum cumulative amplitude is chosen as the filter response. For the second feature, only wavelet coefficients with non-zero ridge indicators are integrated. This procedure allows for variations of the transient to be classified in both time and scale. The sequence c was chosen to maximize the F-ratio of $f_i^{[\text{chirp1}]}$ and $f_i^{[\text{chirp2}]}$ with respect to the training data set discussed in Chapter 8.

If a diagonal or 'chirp-like' time-scale distribution of a transient is found in the scalogram, large values for $f_i^{[\text{chirp1}]}$ and $f_i^{[\text{chirp2}]}$ are obtained with the procedure presented in this subsection. In contrast to a time-domain (FIR) filter the phase of the transient does not need to be defined for the chirp filters.

7.5.2.8 Electrode Artifact Indicators

Electrode artifacts may cause strong, prolonged symmetric waveforms on adjacent channels. Therefore, the squared Euclidean distance between a channel and the inverted adjacent channel

$$f_i^{[\mathrm{dist}]} = \sum_{j=1}^{N_s} \left(v_{ij} + v_{(i+1)j}\right)^2 \ . \tag{7.25}$$

is used as an indicator for electrode artifacts. Focal spikes may lead to similar symmetric waveforms. However, spikes would only last for about 70 ms.

7.5.3 Evaluation of the single-channel features

7.5.3.1 Development Cycle

To establish the single-channel features described the previous subsections a study was performed on a representative training data set of ETs and artifacts. The design involved the repeated evaluation of all 53 EEG recordings in the training set as detailed in section 8.3.1. Features were added to the system

- to identify and verify ETs,
- to optimize the discrimination of ETs from common artifacts, and
- to reject specific artifacts (e.g., electrode or blink artifacts).

After each evaluation of the training set (2-4 h of processing time on a Pentium II, 500 MHz), missed detections and false detections were reviewed. Features were then added, removed or adjusted according to the corresponding F-ratios derived from an *ANOVA*. Thus, the sequence of the chirps filters (Eq. 7.20) was optimized.

7.5.3.2 Bayesian Evaluation of the Discriminant Function

A feature vector is constructed with the single-channel features described in the previous subsections

$$X = \left[s_{i(1)}, s_{i(3)}, s_{i(5)}, s_{i(7)}, s_{i(9)}, m_{i(5)}, m_{i(7)}, m_{i(9)}, f_i^{[\mathrm{chirp1}]}, f_i^{[\mathrm{chirp2}]}, f_i^{[\mathrm{EOG}]}, f_i^{[\mathrm{dist}]}, f_i^{[\mathrm{sharp}]}\right]$$

and used in a linear discriminant function. In a set of 100 EDs marked as definite by two or three readers (see section 8.3.1), 360 ETs were identified on individual channels. A set of feature vectors was compiled for the 360 definite ETs. A second set of 24168 feature vectors was compiled for false detections as found by the Stage 1 detector.

The two sets were considered samples of normal and abnormal classes of EEG activity. Fisher's linear discriminant function $g(X)$ was generated for the two samples. Normalized histograms for the $g(X)$ of both classes are shown in Fig. 7.14. The Bayes

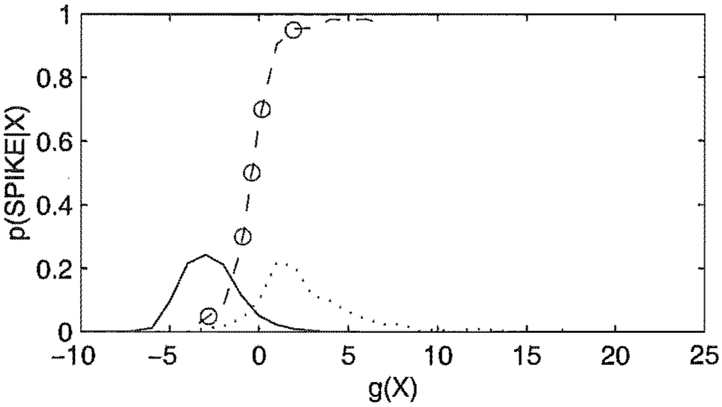

Figure 7.14 Distribution of the single-channel discriminant function $g(X)$ for non-spikes (solid line) and spikes (dotted line). The Bayes probability for having detected a spike given $g(X)$ is indicated by dashed line. Circles mark $p = 0.05$, 0.3, 0.5, 0.7 and 0.95, respectively.

probability for having detected a spike given $g(X)$ is also indicated. Five samples of the Bayes probability were derived from the histogram, indicating the $g(X)$ for detection probabilities p = 0.05, 0.3, 0.5, 0.7 and 0.95 (indicated by circles). The detection probability for each ET was computed by linear interpolation between these five samples.

7.5.4 Multichannel Features

If one or more CETs of an CED are assigned a probability $p > 0.7$, multichannel features are extracted from the blackboard and analyzed. The potential distribution of the CED at the fiducial point (represented by the vector $\mathbf{v}_{\text{obs}}$) is approximated by the real part of the scale-12 wavelet coefficients

$$v_{\text{obs}(i)} = \Re\{w_{iu(3)}\} \tag{7.26}$$

to suppress the influence of coincidental offsets and EMG artifacts.

7.5.4.1 Source-Orientation Estimate

The weighted nodes approach (described in the previous chapter) was used to estimate the source-orientation of $\mathbf{v}_{\text{obs}}$. The resulting source-orientation estimate has one radial and two tangential components

$$\bar{\mathbf{j}} = \begin{pmatrix} \bar{\jmath}_r \\ \bar{\jmath}_s \\ \bar{\jmath}_t \end{pmatrix} . \tag{7.27}$$

The radial component $\bar{\jmath}_r$ was used as a multichannel feature

$$f^{[\mathrm{rad}]} = \bar{\jmath}_r . \tag{7.28}$$

Since $|\bar{\mathbf{j}}| = 1$, $\bar{\jmath}_r$ is equal to the cosine of the angle of the estimated source-orientation with respect to the radial direction. For the weighted nodes approach, the parameter α needs to be chosen as outlined in section 6.7.2. Here, α was optimized to maximize the F-ratio of both classes of transients in the training data set, yielding $\alpha = 0.57$. The distribution of $f^{[\mathrm{rad}]}$ is shown in Fig. 7.15 for non-spikes and spikes. For most spikes the

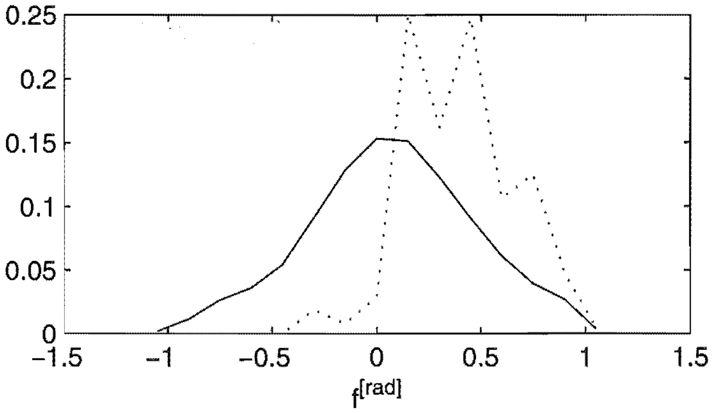

Figure 7.15 Distribution of radial component of orientation estimate $f^{[\mathrm{rad}]}$ for non-spikes and spikes.

source-orientation indicator is positive which corresponds to surface negativity. Spikes with negative orientation indicators correspond to the complex EDs of generalized epilepsy. These EDs show polyphasic ETs that lack a unique fiducial point.

7.5.4.2 Combined Probabilities

The single-channel ET probabilities p_i are combined to form a multichannel feature. The RMS of probabilities is used

$$f^{[\mathrm{prms}]} = \left(\sum_i p_i^2 \right)^{1/2} . \tag{7.29}$$

The feature $f^{[\mathrm{prms}]}$ corresponds to generalized discharges on multiple channels. That is, it is not sensitive to focal events that only involve few channels. The second largest single-channel probability

$$f^{[\mathrm{p2}]} = \max_{i \neq \mathrm{argmax} p_i} p_i \tag{7.30}$$

is also used as a multichannel feature.

7.5.4.3 Adjacency Filter

The potential distribution of focal events shows patterns of phase reversal on adjacent channels which are possibly separated by a null channel. A spatial filter was constructed to capture this spatial feature. The 16 channels were broken down into 2-4 electrode chains depending on the montage used. The adjacency indicator sequences are defined as

$$\rho_i = \begin{cases} 2 - p_i - p_{(i+1)} & \text{if } v_{\mathrm{obs}(i)} < 0 \text{ and } v_{\mathrm{obs}(i+1)} > 0 \text{ and } c_{i(i+1)} > 0 \\ \infty & \text{otherwise} \end{cases} \tag{7.31}$$

with chain indicator

$$c_{ij} = \begin{cases} 1 & \text{if } i, j \text{ are on the same chain} \\ 0 & \text{otherwise,} \end{cases} \tag{7.32}$$

for potentials $v_{\mathrm{obs}(i)}$ and Stage 1 probabilities p_i. If a polarity inversion occurs in an electrode chain and the respective channels feature a high single-channel spike probability, the indicator ρ_i will be small. To cover the possibility of an intervening null channel, a second indicator is defined:

$$\nu_i = \begin{cases} 2 - p_i + p_{(i+1)} - p_{(i+2)} & \text{if } v_{\mathrm{obs}(i)} < 0 \text{ and } v_{\mathrm{obs}(i+2)} > 0 \text{ and } c_{i(i+2)} > 0 \\ \infty & \text{otherwise.} \end{cases} \tag{7.33}$$

The adjacency feature is the overall minimum of all adjacency indicators

$$f^{[\mathrm{adj}]} = \min\{ \min_j \rho_j \, , \, \min_j \nu_j \} \, . \tag{7.34}$$

7.5.5 Final Classification

For each CED, the multichannel feature vector

$$X' = [f^{[\mathrm{rad}]}, f^{[\mathrm{prms}]}, f^{[\mathrm{p2}]}, f^{[\mathrm{adj}]}]$$

is constructed. The set of 100 definite EDs (same as used in section 7.5.3.2) and the set of all 24168 non-spike CEDs detected by Stage 1 were used to obtain the multichannel linear discriminant function $g'(X')$. The distributions for spikes and non-spikes are

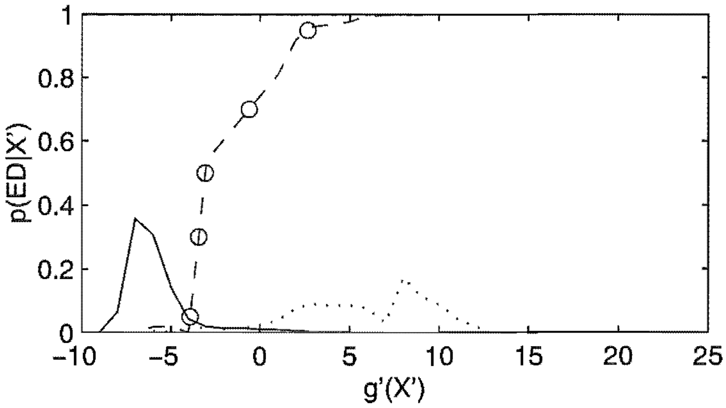

Figure 7.16 Distributions of the multichannel discriminant $g'(X')$ for non-spikes (solid line) and spikes (dotted line). The Bayes probability for a spike detection given $g'(X')$ is indicated by dashed line. Dots mark $p = 0.05,\ 0.3,\ 0.5,\ 0.7$ and 0.95, respectively.

shown in Fig. 7.16. The separation between the two classes is substantially larger than for the single-channel discriminant. A CED is classified as an ED if

$$g'(X') > 0.7 \tag{7.35}$$

but focal CEDs can be upgraded to a focal ED for smaller values of $g'(X')$ if the condition

$$\begin{array}{llll} & g'(X') > 0.3 & & \\ \text{and} & \max\{p_i, p_{(i+\mu)}\} > 0.8 & \text{and} & \min\{p_i, p_{(i+\mu)}\} > 0.6 \\ \text{and} & v_{\mathrm{obs}(i)} < 0 & \text{and} & v_{\mathrm{obs}(i+\mu)} > 0 \\ \text{and} & c_{i(i+\mu)} > 0 & & \end{array} \tag{7.36}$$

is met for any channel i and $\mu = 1, 2$. In other words, if two adjacent channels or two channels separated by only one channel exhibit high spike probabilities and phase reversal, a focal ED is deemed to have been detected. Detection thresholds in equations 7.35 and 7.36 were derived from the analysis of the training set. The criterion used was that at least two definite EDs should be detected in each recording known to contain definite EDs. Classification results for the training set and the test set are provided in detail and discussed in the next chapter.

7.6 SUMMARY

A two-stage wavelet-based spike detection system has been proposed. The Stage 1 transient detector includes rejection mechanisms for EMG and EOG activity. The multichannel multiscale context of a CED is computed and analyzed in Stage 2. Thirteen single-channel features are extracted. A linear discriminant function is obtained from the analysis of a training data set. The output of a linear discriminant function is interpreted with a Bayesian approach, yielding a CET probability for each channel. Four multichannel features are derived from the spatial distribution of the potential and the single-channel probabilities. A second linear discriminant function is employed to obtain the final spike probability for each multichannel transient. An additional rule based on adjacency patterns is used to upgrade focal events with lower spike probabilities.

Chapter 8

SYSTEM EVALUATION

8.1 INTRODUCTION

To assess the performance of the wavelet-based spike detection system presented in the previous chapter, the system was tested extensively. The data set from which the training and test data sets for the system were derived is presented. Since events in the data set were marked as either 'questionable' or 'definite', a brief discussion of the problem of incorporating such perception values in the sensitivity formula is presented.

For each recording in both the training and test sets the number of missed definite EDs and the number of false detections were determined. Based on these numbers, sensitivities, selectivities and false detection rates are given for each recording and as an average for the training and test sets.

The results are compared with those obtained with two hybrid spike detection systems on the same data set: the system developed by Dingle *et al.* [1993], based on a mimetic stage and a rule-based expert system, and the system developed by James *et al.* [1999] based on mimetic, ANN and fuzzy logic stages.

In the evaluation of the spike detection systems several characteristic missed definite EDs and false detections are listed and discussed. Figures of the respective events are found in Appendix A.

Performance of the three systems is also compared at the 'global EEG level' in which classification is in terms of whether the entire recordings contain epileptiform activity or not. Several performance studies reported in the literature are presented as a reference.

8.2 PERFORMANCE MEASURES

8.2.1 Traditional Measures

Traditional performance measure such as sensitivity, selectivity and the false detection rate were introduced in section 5.3.1. They relate to classifications on a *dichotomous* *scale*, that is events are either EDs or artifacts, but no intermediate states or class memberships are possible.

In this study, both clinical experts and spike detection systems assigned *perception values* such as 'questionable', 'possible', 'probable' or 'definite' to an ED.

Unfortunately, there is no 'gold standard' for the performance evaluation of a spike detection system, since EEGers disagree on a considerable number of events. Therefore, events not marked as definite by at least two of the three readers were arbitrarily excluded from the evaluation. Such events were neither counted as false detections if picked up by the system nor counted as missed detections if missed by the system.

Sensitivity was measured by the ratio of definite EDs found by the system to the total number of definite EDs (Eq. 5.10). Selectivity was measured by the ratio of definite EDs found by the system to the total number of EDs found by the system (Eq. 5.11). False detections were counted if none of the readers had marked the respective event[1].

8.2.2 Continuous-Valued Sensitivities and Selectivities

Conventional performance measures such as sensitivity and selectivity are based on counts of detections. Thus, to apply these measures each CED needs to be evaluated as either being detected or not detected in what is a *dichotomous scale*. Wilson *et al.* [1996] tried to utilize the additional information given by gradual or continuous perception values other than 'definite' or 'none' in a new formula for the sensitivity and selectivity. They suggested the following formula for the continuous-valued sensitivity of reader A with respect to reader B

$$Sensitivity_{\mathrm{AB}} = \frac{\sum_{i=1}^{N^{\mathrm{A}\cup\mathrm{B}}} x_{\mathrm{A}i} \cdot x_{\mathrm{B}i}}{\sum_{i=1}^{N^{\mathrm{A}\cup\mathrm{B}}} x_{\mathrm{B}i}^2} = Selectivity_{\mathrm{BA}} \tag{8.1}$$

where $N^{\mathrm{A}\cup\mathrm{B}}$ is the number of events reported by one or both readers and $x_{\mathrm{A}i}$ is the perception value given by reader A to event i. Each perception value $x_{\mathrm{A}i}$ can be interpreted as a subjective probability assigned by a reader to an event. If a reader

Table 8.1 Perception values $x_{\mathrm{X}i}$ given to event i by reader X [Black and Jones 1998].

	Readers		
	A	B	C
Event 1	1.000	1.000	0.750
Event 2	0.000	0.250	0.500
Event 3	0.250	0.000	0.000
Event 4	0.500	0.250	0.000

did not report an event the respective perception value is assumed to be zero. Black

[1]Interestingly, some of these 'false detections' were reviewed by one of the original 3 EEGers at a later stage and subsequently noted as 'questionable' or 'definite'.

and Jones [1998] pointed out that a sensitivity reading larger than 1.0 can result from Eq. 8.1 making it inappropriate to interpret the result of Eq. 8.1 from the viewpoint of traditional understandings of sensitivity or selectivity [Black *et al.* 1997].

An improvement to Wilson's continuous-valued sensitivity can be achieved by an alternative definition

$$Sensitivity_{\mathrm{AB}} = \left(\frac{\sum_{i=1}^{N^{\mathrm{A}\cup\mathrm{B}}} \sqrt{x_{\mathrm{A}i} \cdot x_{\mathrm{B}i}}}{\sum_{i=1}^{N^{\mathrm{A}\cup\mathrm{B}}} x_{\mathrm{B}i}} \right)^2 = Selectivity_{\mathrm{BA}} \ . \tag{8.2}$$

In an example given by Black and Jones [1998] four events are rated by three readers as listed in Table 8.1. When inter-reader sensitivities are evaluated according to Eq. 8.1

Table 8.2 Inter-reader sensitivities and selectivities according to Eq. 8.1 [Black and Jones 1998]. The sensitivity of reader C with respect to reader B is found in the second line of the third column.

Reader	A	B	C	Average selectivity
A	1.000	0.857	0.571	0.714
B	1.000	1.000	0.778	0.889
C	0.923	1.077	1.000	1.000
Average sensitivity	0.962	0.967	0.675	0.868

Table 8.3 Inter-reader sensitivities and selectivities according to Eq. 8.2.

Reader	A	B	C	Average selectivity
A	1.000	0.598	0.245	0.614
B	0.814	1.000	0.661	0.825
C	0.480	0.952	1.000	0.811
Average sensitivity	0.765	0.850	0.635	0.750

(Table 8.2), reader A is assigned a sensitivity of 100% with respect to reader B event though reader A missed event 2. The evaluation according to Eq. 8.2 is listed in Table 8.3. In this, a sensitivity of 81.4% is estimated for reader A with respect to reader B giving a more appropriate reflection of the situation. However, this continuous-valued formula is also not confined to a range of 0.0 to 1.0, and, hence, the traditional method based on a *dichotomous scale* will be used in the following evaluations.

8.3 EEG DATA SET

Routine EEG recordings were obtained from 521 consecutive patients referred to the Neurology Department at Christchurch Hospital. The patients' ages ranged from 2 weeks to 86 years. Seven bipolar and referential montages were used in the recordings. Recordings were made while the patient was awake but resting. Periods of eyes open, eyes closed, hyperventilation and photic stimulation were included. Scalp potentials were recorded in 16 bipolar channels derived from 21 electrodes placed according to the International 10-20 system (Fig. 1.5) [Jasper 1958]. Channels were amplified and recorded with a Siemens Minograph Universal EEG machine after bandpass-filtering between 0.5 to 70 Hz (5-pole Butterworth filter). Simultaneously, the EEG was sampled at 200 Hz, digitized with 12-bit resolution and stored for off-line processing.

Recordings were re-referenced off-line to four montages including longitudinal, transversal, longitudinal/transverse and circumferential as illustrated in Fig. 6.1. Thus, bipolar montages were obtained throughout the entire recording.

Once the clinical neurophysiologist in charge had identified 50 definitely epileptiform recordings the data collection was stopped. The set of 521 recordings has been evaluated in a previous study by Black *et al.* [2000].

8.3.1 Data Subsets for Training and Testing

Of the 521 recordings, 106 were evaluated by three independent EEG readers, the remaining 415 were seen by two EEGers. The former subset was used in this study. Of the 106, 17 recordings contained epileptiform activity according to at least two of the three EEG readers with one reader only reporting epileptiform activity in a further 10 recordings. 13 recordings contained epileptiform activity according to all three EEGers[2]. The remaining 79 recordings were considered to contain no epileptiform activity. In the subset, the patients' age ranged from 2 weeks to 84 years (mean 35.0 years), the length of the recordings ranged from 12 min 28 s to 28 min 23 s (mean 23 min 44 s). The readers categorized the EA as being generalized in 28 EEGs, focal in 13 EEGs, multifocal in 10 EEGs, and lateralized in 2 EEGs as listed in the third column of tables 8.5 and 8.6.

The subset of 106 was split up into a training set (13 epileptiform, 40 normal, duration: 21 h 20 min) and a test set (14 epileptiform, 39 normal, duration: 20 h 37 min). The training set was used in the design of the system, specifically for the training of the linear discriminant functions. Of the 13 epileptiform recordings in the training set, 6 contained definite EDs used for actual training of the system. The consensus between EEGers was considered insufficient for the events in the remaining 7 recordings.

8.3.2 Recording Protocols

In order to capture possible interictal and/or ictal EEG activity, clinical recordings were carried out according to one of 3 standard protocols, selected according to the subject's age: 'BABY' protocol for subjects under 12 months, 'UNDER5' protocol for subjects between the age of 1 and 5, and 'OVER5' protocol for subjects aged 5 or older

[2]This emphasizes the level of disagreement between EEGers, even at the global EEG level, and highlights the difficulty in using EEGers as the gold standard against which the performance of the spike detection system is measured.

Table 8.4 Protocols used in clinical recordings. Some runs were re-referenced in the digital EEG to obtain bipolar recordings throughout.

Run	Montage on paper	Montage in file	Stimuli/ Action	Duration
		Protocol BABY		
1	Longitudinal	Longitudinal	none	500 s
2	Longitudinal	Longitudinal	Photic stimulation rising from 5 to 30 Hz	60 s
		Protocol UNDER5		
1	Longitudinal	Longitudinal	none	200 s
2	Transverse	Transverse	none	100 s
3	Longitudinal-transverse	Longitudinal-transverse	none	100 s
4	Circumferential	Circumferential	none	100 s
5	Average reference	Longitudinal	none	100 s
6	Vertex reference	Longitudinal	none	100 s
7	Longitudinal	Longitudinal	Photic stimulation rising from 5 to 30 Hz	60 s
		Protocol OVER5		
1	Longitudinal	Longitudinal	none	200 s
2	Ipsilateral ears reference	Longitudinal	none	100 s
3	Transverse	Transverse	none	100 s
4	Longitudinal-transverse	Longitudinal-transverse	none	100 s
5	Circumferential	Circumferential	none	100 s
6	Average reference	Longitudinal	none	100 s
7	Ipsilateral ears reference	Longitudinal	none	100 s
8	Vertex reference	Longitudinal	none	100 s
9	Longitudinal	Longitudinal	Hyperventilation	300 s
10	Longitudinal	Longitudinal	Photic stimulation rising from 5 to 30 Hz	100 s

(Table 8.4). Hyperventilation and photic stimulation with a stroboscopic flash were used to increase likelihood of epileptic seizures during the recording. The montages used in the digital recordings are presented in detail in Chapter 6 (p. 84).

8.4 TRAINING OF THE WAVELET-BASED SPIKE DETECTION SYSTEM

8.4.1 Refinement of the Training Data Set

To utilize the evaluations of the three independent EEGers (one clinical neurophysiologist and two neurologists) for training and testing of the wavelet-based system, only EDs marked as definite by two or three readers were regarded as *definite EDs* and used for training. 200 definite EDs in the training set were picked up by stage 1 of the wavelet-based system. Of these, 100 EDs which showed clear artifact-free ETs on several channels were chosen. Finally, from this subset 360 ETs were selected on individual channels that clearly showed an epileptiform discharge.

8.4.2 Training of the Linear Discriminant Functions

A large number of 24168 artifacts (events not marked as questionable or definite by any EEGer) were detected as CEDs by stage 1. Only the channel showing the largest amplitude was selected in each case to represent an artifactual single-channel transient. Thus, 24168 artifactual single-channel transients and 360 ETs were used to train the single-channel linear discriminant function[3].

The system was run twice on the training data set to complete the training. In the first training run, all 13 single-channel features were obtained for both classes of single-channel transients, yielding a 360×13 matrix for the ETs and a 24168×13 matrix for the artifacts. Fisher's linear discriminant function was obtained for both statistical samples.

Two histograms were generated to estimate the Bayesian probability for having detected a spike given the output of the LDF. Five samples were taken of the Bayesian probability for having detected a spike (Fig. 7.14).

In the second training run, the single-channel LDF was used to obtain the multichannel features of each ED and each artifact. Again, the Bayesian probability was estimated from histograms (Fig. 7.16). Altogether, 24168 artifacts and 100 EDs were used to train the multichannel LDF.

8.5 PERFORMANCE OF THE WAVELET BASED SPIKE DETECTION SYSTEM

The trained system has been run on both the training and the test sets. Only definite EDs were counted in the evaluation of the system performance. Sensitivities, selectivities and false detection rates of the training set and test set are listed in Tables 8.5 and 8.6, respectively.

8.5.1 Training Data Set

The mean performance on the training set was 64.7% sensitivity, 67.8% selectivity, and an average 8.0 false detections per hour. On average there were 17.5 false detections per hour in the 13 epileptiform EEGs and 4.9 false detections per hour in the 40 normal controls. For the three EEGs showing generalized epileptiform activity the mean performance was 97% sensitivity and 84% selectivity. In contrast, for the three focal and multifocal EEGs, the mean performance dropped to 32% sensitivity and 67% selectivity.

[3]The massive imbalance between the number of EDs and the number of artifacts was compensated by the use of *normalized* histograms for the estimation of Bayesian probabilities in section 5.5.

8.5.2 Test Data Set

For the test set (Table 8.6), the mean performance was lower with 51.0% sensitivity, 46.8% selectivity but only 5.6 false detections per hour on average. 58 of the 105 false detections were in the 40 controls; the mean false detection rate was 5.9 for the 14 epileptiform EEGs and 4.1 in the 39 controls. No false detections were made in 21 of the 40 controls. There were more than 2 false detections in only 6 of the normal EEGs.

8.5.3 Missed Detections

Many of the EDs in EEG 6 (E8434) were characterized by pairs of focal EDs 300 ms apart. In many of the pairs, the second ED was detected since it had a larger amplitude than the first one. An EEGer can establish a focus following the detection of the more prominent EDs throughout the recording. Subsequently, possible EDs with smaller amplitudes but the same spatial distribution as the definite EDs could be upgraded to definite EDs. Thus, the wide-sense temporal context is evaluated. In EEG 6, this evaluation is facilitated by the close temporal proximity of several EDs. The resulting sensitivity of the system was 28% (Fig. A.16 on p. 171).

Only 19% of the focal EDs in EEG 12 (E8219) were picked up by the wavelet-based system (Fig. A.6 on p. 161). Many of the definite EDs had moderately high single-channel probabilities ($p = 0.5 \ldots 0.75$) but were not upgraded due to not meeting focal upgrade criteria (Eq. 7.36).

All four definite focal EDs in EEG 10 (E2289B) occurred in runs which were presented in referential montage to the readers but were processed in longitudinal montage by the system. Although the EDs are prominent on paper (i.e., referential montage), they effectively substantially attenuated in the re-referencing process. This notwithstanding, the detection system successfully detected all 5 definite EDs (p. 157).

Only 5 (9%) of the 56 definite EDs in EEG 9 (E8557) were detected. All of these spikes had an uncharacteristic increasing slope prior to the spike which resulted in low single-channel probabilities (Fig. A.20 on p. 175). This could be due to the young age of the patient (2 weeks, youngest in the data set). It should be noted that the inter-reader sensitivities and selectivities for this recording ranged from 9.9 to 64%.

All 15 definite EDs in EEG 11 (E8159B) were missed by the system because they featured sharp waves without spikes. This variant did not occur in the training set (Fig. A.4 on p. 159).

8.5.4 False Detections

In EEG 22 (E8559) a combination of electrode and movement artifacts triggered 29 false detections which were unrelated to the EDs marked by one reader in the same recording (Fig. A.21 on p. 176).

Table 8.5 Performance of the **wavelet-based** spike detection system on the **training data set**. The type of epileptiform activity as diagnosed by the three readers is indicated by G (generalized), F (focal), M (multifocal), L (lateralized) or N (none). Also listed are the number of definite EDs (Def), the number of missed definite EDs (Mis), the number of false detections (Fal), the sensitivity (Sen), the selectivity (Sel) and the number of false detections per hour (FD/hr).

Recording				Stage 1				Stage 2				
EEG	Age	Type	Def	Mis	Fal	Sen	Sel	Mis	Fal	Sen	Sel	FD/hr
2 E3394B	12 years	GGG	22	0	522	100%	4%	2	1	91%	95%	2.3
4 E2104C	16 years	GGG	3	0	507	100%	1%	0	0	100%	100%	0.0
6 E8434	8 years	MFM	189	53	271	72%	33%	136	17	28%	76%	38.2
8 E2289B	84 years	MNF	4	0	172	100%	2%	2	5	50%	29%	11.2
10 E8372	5 years	GGG	5	0	582	100%	1%	0	12	100%	29%	26.3
12 E8219	4 years	MMM	36	6	246	83%	11%	29	2	19%	78%	8.0
14 E8438	37 years	NFF	-	-	372	-	-	-	3	-	-	7.3
16 E8263	25 years	GGN	-	-	517	-	-	-	0	-	-	0.0
18 E8349	49 years	NFN	-	-	415	-	-	-	6	-	-	13.5
20 E2289C	84 years	NFN	-	-	224	-	-	-	1	-	-	2.5
22 E8559	3 years	NFN	-	-	638	-	-	-	29	-	-	111.5
24 E7421B	29 years	FNN	-	-	463	-	-	-	0	-	-	0.0
26 E8288	20 years	NGN	-	-	462	-	-	-	3	-	-	6.6
28 E8344	57 years	NNN	-	-	581	-	-	-	2	-	-	4.8
30 E8276	42 years	NNN	-	-	494	-	-	-	0	-	-	0.0
32 E8297	6 months	NNN	-	-	730	-	-	-	2	-	-	8.8
34 E3507D	29 years	NNN	-	-	230	-	-	-	1	-	-	2.4
36 E8443	25 years	NNN	-	-	1391	-	-	-	6	-	-	16.5
38 E1009B	35 years	NNN	-	-	293	-	-	-	1	-	-	2.7
40 E8380	7 months	NNN	-	-	584	-	-	-	15	-	-	50.6
42 E8419	43 years	NNN	-	-	529	-	-	-	12	-	-	27.4
44 E8585	14 years	NNN	-	-	468	-	-	-	0	-	-	0.0
46 E8600	14 years	NNN	-	-	395	-	-	-	3	-	-	7.3
48 E8260	65 years	NNN	-	-	304	-	-	-	2	-	-	4.2
50 E8545	25 years	NNN	-	-	575	-	-	-	0	-	-	0.0
52 E8331	32 years	NNN	-	-	530	-	-	-	2	-	-	4.4
54 E8625	62 years	NNN	-	-	572	-	-	-	1	-	-	2.1
56 E8572	32 years	NNN	-	-	386	-	-	-	0	-	-	0.0
58 E8580	65 years	NNN	-	-	307	-	-	-	1	-	-	2.4
60 E8595	26 years	NNN	-	-	668	-	-	-	0	-	-	0.0
62 E8317	14 years	NNN	-	-	352	-	-	-	0	-	-	0.0
64 E8424	73 years	NNN	-	-	180	-	-	-	1	-	-	2.3
66 E8458	13 years	NNN	-	-	427	-	-	-	0	-	-	0.0
68 E8611	8 weeks	NNN	-	-	106	-	-	-	0	-	-	0.0
70 E8615	30 years	NNN	-	-	332	-	-	-	13	-	-	31.0
72 E8622	62 years	NNN	-	-	634	-	-	-	0	-	-	0.0
74 E8324	77 years	NNN	-	-	127	-	-	-	0	-	-	0.0
76 E8375	43 years	NNN	-	-	319	-	-	-	0	-	-	0.0
78 E8237	62 years	NNN	-	-	537	-	-	-	1	-	-	2.4
80 E8407	48 years	NNN	-	-	710	-	-	-	1	-	-	2.4
82 E8252	73 years	NNN	-	-	135	-	-	-	0	-	-	0.0
84 E8453	48 years	NNN	-	-	1135	-	-	-	0	-	-	0.0
86 E4187B	8 years	NNN	-	-	274	-	-	-	0	-	-	0.0
88 E8365	40 years	NNN	-	-	395	-	-	-	1	-	-	2.3
90 E8304	20 years	NNN	-	-	455	-	-	-	1	-	-	2.3
92 E8200	60 years	NNN	-	-	455	-	-	-	0	-	-	0.0
94 E8232	12 years	NNN	-	-	198	-	-	-	3	-	-	6.9
96 E8446	67 years	NNN	-	-	544	-	-	-	3	-	-	9.3
98 E8528	21 years	NNN	-	-	953	-	-	-	0	-	-	0.0
100 E8387	35 years	NNN	-	-	296	-	-	-	1	-	-	2.4
102 E8563	16 years	NNN	-	-	500	-	-	-	0	-	-	0.0
104 E8556	27 years	NNN	-	-	461	-	-	-	1	-	-	2.3
106 E8462	65 years	NNN	-	-	215	-	-	-	0	-	-	0.0
mean(sum)	35.0 years		(259)	(59)	(24168)	92.5%	8.7%	(169)	(153)	64.7%	67.8%	8.0

Table 8.6 Performance of the **wavelet-based** spike detection system on the **test data set.**

Recording				Stage 1				Stage 2				
EEG	Age	Type	Def	Mis	Fal	Sen	Sel	Mis	Fal	Sen	Sel	FD/hr
1 E1109D	46 years	GGG	50	1	649	98%	46%	18	0	64%	100%	0.0
3 E776C	17 years	GGG	2	0	365	100%	8%	0	2	100%	90%	4.7
5 E8223	50 years	GGN	1	0	476	100%	30%	0	10	100%	62%	28.3
7 E8247	56 years	GGN	3	0	282	100%	3%	2	1	33%	67%	2.5
9 E8557	2 weeks	MMM	56	3	1043	95%	17%	51	2	9%	78%	7.5
11 E8159B	65 years	GGL	15	0	334	100%	25%	15	2	0%	0%	5.1
13 E8188B	7 months	GLF	-	-	531	-	-	-	4	-	-	17.0
15 E8314	33 years	NGG	-	-	374	-	-	-	2	-	-	4.7
17 E8361	25 years	MFN	-	-	567	-	-	-	1	-	-	2.3
19 E8293	2 years	NFN	-	-	441	-	-	-	2	-	-	7.9
21 E8242	2 years	NGN	-	-	232	-	-	-	2	-	-	6.5
23 E8214	62 years	GNN	-	-	407	-	-	-	10	-	-	24.5
25 E8618	68 years	FNN	-	-	347	-	-	-	0	-	-	0.0
27 E8396	70 years	NNF	-	-	550	-	-	-	9	-	-	21.9
29 E8523	30 years	NNN	-	-	418	-	-	-	2	-	-	4.9
31 E8399	48 years	NNN	-	-	376	-	-	-	5	-	-	16.4
33 E8268	10 years	NNN	-	-	595	-	-	-	2	-	-	4.9
35 E4774B	54 years	NNN	-	-	398	-	-	-	0	-	-	0.0
37 E8507	61 years	NNN	-	-	348	-	-	-	0	-	-	0.0
39 E8416	26 years	NNN	-	-	1087	-	-	-	4	-	-	9.2
41 E8514	6 years	NNN	-	-	263	-	-	-	7	-	-	18.0
43 E8383	26 years	NNN	-	-	587	-	-	-	1	-	-	2.4
45 E8490	63 years	NNN	-	-	892	-	-	-	0	-	-	0.0
47 E8392	14 months	NNN	-	-	515	-	-	-	1	-	-	3.9
49 E8518	13 years	NNN	-	-	558	-	-	-	7	-	-	19.6
51 E8449	7 years	NNN	-	-	362	-	-	-	1	-	-	2.4
53 E8627	61 years	NNN	-	-	349	-	-	-	0	-	-	0.0
55 E8485	13 years	NNN	-	-	480	-	-	-	1	-	-	2.3
57 E8308	79 years	NNN	-	-	222	-	-	-	2	-	-	4.6
59 E8579B	71 years	NNN	-	-	256	-	-	-	1	-	-	3.1
61 E8568	16 years	NNN	-	-	612	-	-	-	0	-	-	0.0
63 E8336	58 years	NNN	-	-	173	-	-	-	0	-	-	0.0
65 E8510	80 years	NNN	-	-	149	-	-	-	0	-	-	0.0
67 E8536	74 years	NNN	-	-	246	-	-	-	0	-	-	0.0
69 E8466	16 years	NNN	-	-	390	-	-	-	0	-	-	0.0
71 E8503	8 years	NNN	-	-	547	-	-	-	0	-	-	0.0
73 E8285	13 years	NNN	-	-	364	-	-	-	0	-	-	0.0
75 E8473	55 years	NNN	-	-	1012	-	-	-	0	-	-	0.0
77 E8272	28 years	NNN	-	-	308	-	-	-	0	-	-	0.0
79 E8233	28 years	NNN	-	-	589	-	-	-	2	-	-	4.8
81 E8412	5 years	NNN	-	-	357	-	-	-	2	-	-	4.6
83 E8112B	39 years	NNN	-	-	487	-	-	-	0	-	-	0.0
85 E8429	78 years	NNN	-	-	587	-	-	-	0	-	-	0.0
87 E8260B	65 years	NNN	-	-	141	-	-	-	2	-	-	4.8
89 E8280	54 years	NNN	-	-	284	-	-	-	0	-	-	0.0
91 E7743B	5 years	NNN	-	-	426	-	-	-	3	-	-	6.7
93 E8575	6 weeks	NNN	-	-	383	-	-	-	0	-	-	0.0
95 E4984B	23 years	NNN	-	-	365	-	-	-	0	-	-	0.0
97 E8478	12 months	NNN	-	-	456	-	-	-	13	-	-	48.1
99 E8541	66 years	NNN	-	-	259	-	-	-	0	-	-	0.0
101 E8356	3 weeks	NNN	-	-	212	-	-	-	0	-	-	0.0
103 E8549	15 years	NNN	-	-	353	-	-	-	0	-	-	0.0
105 E8605	64 years	NNN	-	-	373	-	-	-	2	-	-	4.8
mean(sum)	35.0 years		(127)	(4)	(23419)	98.8%	21.5%	(86)	(105)	51.0%	46.8%	5.6

A high false detection rate (15 events, 50.6/h) was found in EEG 40 (E8380). The EEG shows sharp focal activity that was not marked as epileptiform by any of the readers (Fig. A.13 on p. 168).

Sharp alpha activity caused 12 false detections in EEG 42 (E8419, Fig. A.15 on p. 170) and 13 false detections in EEG 70 (E8615, Fig. A.22 on p. 177).

Sharp focal activity caused 10 false detections in EEG 23, which could be partly due to reformatting (E8214, Fig. A.5 on p. 160). Complex electrode artifacts caused 13 false detections in EEG 97 (E8478, Fig. A.18 on p. 173).

8.6 COMPARISON WITH TWO OTHER SYSTEMS

The entire data set of 106 EEGs was evaluated by two other spike detection systems, Hybrid I and Hybrid II, previously developed at Christchurch Hospital.

8.6.1 Hybrid I System

The Hybrid I system features two stages: mimetic feature extraction and rule-based evaluation [Dingle 1992, Dingle *et al.* 1993, Jones *et al.* 1994, Jones *et al.* 1996, Black *et al.* 2000]. Both stages were evaluated separately. A large number of single-channel events are put forward by the mimetic feature extraction stage. To measure the detection properties of stage one, single-channel events were grouped into clusters if they occurred less than 125 ms apart. These clusters were counted as stage one detections. Stage two detections were rated as possible, probable or definite by the rule-based system. To obtain maximum sensitivity all detections were considered here. A more detailed evaluation of the three perception values is presented in section 8.7. The results for each recording in the test and training data sets are listed in Tables 8.7 and 8.8, respectively.

The system failed to find any of the definite events in 5 recordings (5, 7, 8, 10, and 12). However, the system detected other events in four of the recordings. Only in recording 8 (E2289B) were all four re-referenced definite events missed (Fig. A.2 on p. 157). The only definite event in recording 5 (E8223) was missed but a questionable event was reported (Fig. A.7 on p. 162). In recording 7 (E8247) three definite events were missed but one event was picked up and reported as possible (Fig. A.8 on p. 163). In recording 10 (E8372) all five definite events were missed but two artifacts were detected (Fig. A.12 on p. 167). In recording 12 (E8219) 36 definite events were missed (p. 161), and a questionable ED and an artifact were detected and ranked 'probable'.

In five of the normal controls (29, 30, 31, 32, and 33) single definite events were incorrectly reported. In recording 29 (E8523) a short eye-blink with sharp biphasic onset is classified as definite nonfocal ED (Fig. A.19 on p. 174). Similar eye-blinks triggered definite nonfocal detections in recordings 30 (E8276) and 31 (E8268).

In recording 31 (E8399, Fig. A.14 on p. 169) the system reported a definite non-focal ED which was confirmed as questionable by a neurophysiologist (GC) on second consultation. An event reported as definite focal ED by the Hybrid I system and the wavelet-based system in recording 32 (E8297, Fig. A.10 on p. 165) was also labelled as questionable on second consultation.

The Hybrid I system is the only system tested in this work which makes use of the wide-sense temporal context of CEDs. It evaluates focal and non-focal CEDs and tries to establish the focus (or foci) of focal events. The importance of this was recently emphasized by Ramabhadran *et al.* [1999]. An advantage of the rule-based evaluation or reasoning is that the cause for any misclassification can be traced back through the system. Just one additional rule may be put in place to cater for a new variety of spikes or artifacts. A potential advantage over statistical classifiers is that the knowledge accumulated by the system is explicitly readable as a set of rules. In contrast, the 'knowledge' of statistical classifiers is represented by a set of weight vectors.

8.6.2 Hybrid II System

The second system, Hybrid II, was also based on mimetic feature extraction followed by a single-channel neural network classifier (SOFM) and multi-channel fuzzy rule-based spatial combiner [James *et al.* 1996a, James 1997, James *et al.* 1999]. Version 2.00 of the system was available for evaluation. Seven of the 106 EEGs had been used in the training of the SOFM (EEGs 1, 2, 3, 4, 6, 8, and 13). Three of those are part of the current test set. The following options were chosen for the Hybrid II system

- a 20 × 20-node SOFM which was fine-tuned with *learning vector quantization* method LVQ2 (file: `lvq1_2.som`),
- a set of 127 fuzzy rules for the spatial combiner (file: `rules_04.fuz`), and
- data driven calculation of the gain value (menu option: 'Calculate Gain Term').

Stage one detections were not available for evaluation. Since the mimetic feature extraction was designed in the same way as in the Hybrid I system, stage one detections can be assumed to be similar to those of the Hybrid I system. Detections were also rated as possible, probable or definite by the system. In contrast to Hybrid I there was a high number of possible detections (> 500) for many of the normal EEGs. Therefore, only probable and definite detections were counted in this evaluation. The system failed to process EEGs 15, 16, 20, 36, and 61 (system crashed). Results for training and test data sets are listed in Tables 8.9 and 8.10, respectively. None of the definite EDs were picked up in EEGs 5, 6, 8, 9, and 12. No false detections were reported in 36 of the 79 normal controls. Large false detection rates (> 50/h) were found in 16 EEGs (7 epileptiform, 9 normal controls). Overall, the Hybrid II system had a mean sensitivity

Table 8.7 Performance of the **Hybrid I** spike detection system (mimetic and rule based) on the **training data set.**

Recording				Stage 1 (Mimetic)				Stage 2 (Mimetic+Rules)				
EEG	Age	Type	Def	Mis	Fal	Sen	Sel	Mis	Fal	Sen	Sel	FD/hr
2	12 years	GGG	22	1	550	95%	4%	11	0	50%	100%	0.0
4	16 years	GGG	3	0	667	100%	0%	2	0	33%	100%	0.0
6	8 years	MFM	189	47	625	75%	19%	149	2	21%	95%	4.5
8	84 years	MNF	4	2	1312	50%	0%	4	0	0%	0%	0.0
10	5 years	GGG	5	1	1295	80%	0%	5	2	0%	0%	4.4
12	4 years	MMM	36	4	260	89%	11%	36	0	0%	0%	0.0
14	37 years	NFF	-	-	967	-	-	-	0	-	-	0.0
16	25 years	GGN	-	-	911	-	-	-	0	-	-	0.0
18	49 years	NFN	-	-	676	-	-	-	0	-	-	0.0
20	84 years	NFN	-	-	2073	-	-	-	0	-	-	0.0
22	3 years	NFN	-	-	738	-	-	-	1	-	-	3.8
24	29 years	FNN	-	-	812	-	-	-	0	-	-	0.0
26	20 years	NGN	-	-	1038	-	-	-	2	-	-	4.4
28	57 years	NNN	-	-	921	-	-	-	2	-	-	4.8
30	42 years	NNN	-	-	688	-	-	-	1	-	-	2.3
32	6 months	NNN	-	-	288	-	-	-	1	-	-	4.4
34	29 years	NNN	-	-	631	-	-	-	1	-	-	2.4
36	25 years	NNN	-	-	531	-	-	-	1	-	-	2.8
38	35 years	NNN	-	-	310	-	-	-	3	-	-	8.2
40	7 months	NNN	-	-	847	-	-	-	2	-	-	6.7
42	43 years	NNN	-	-	599	-	-	-	1	-	-	2.3
44	14 years	NNN	-	-	837	-	-	-	1	-	-	2.3
46	14 years	NNN	-	-	780	-	-	-	1	-	-	2.4
48	65 years	NNN	-	-	1232	-	-	-	1	-	-	2.1
50	25 years	NNN	-	-	899	-	-	-	1	-	-	2.4
52	32 years	NNN	-	-	1297	-	-	-	1	-	-	2.2
54	62 years	NNN	-	-	1958	-	-	-	0	-	-	0.0
56	32 years	NNN	-	-	1159	-	-	-	0	-	-	0.0
58	65 years	NNN	-	-	1511	-	-	-	0	-	-	0.0
60	26 years	NNN	-	-	852	-	-	-	0	-	-	0.0
62	14 years	NNN	-	-	598	-	-	-	0	-	-	0.0
64	73 years	NNN	-	-	1330	-	-	-	0	-	-	0.0
66	13 years	NNN	-	-	826	-	-	-	0	-	-	0.0
68	8 weeks	NNN	-	-	54	-	-	-	0	-	-	0.0
70	30 years	NNN	-	-	666	-	-	-	0	-	-	0.0
72	62 years	NNN	-	-	785	-	-	-	0	-	-	0.0
74	77 years	NNN	-	-	1818	-	-	-	0	-	-	0.0
76	43 years	NNN	-	-	1013	-	-	-	0	-	-	0.0
78	62 years	NNN	-	-	399	-	-	-	0	-	-	0.0
80	48 years	NNN	-	-	1784	-	-	-	0	-	-	0.0
82	73 years	NNN	-	-	1109	-	-	-	0	-	-	0.0
84	48 years	NNN	-	-	1005	-	-	-	0	-	-	0.0
86	8 years	NNN	-	-	913	-	-	-	0	-	-	0.0
88	40 years	NNN	-	-	1008	-	-	-	0	-	-	0.0
90	20 years	NNN	-	-	566	-	-	-	0	-	-	0.0
92	60 years	NNN	-	-	504	-	-	-	0	-	-	0.0
94	12 years	NNN	-	-	399	-	-	-	0	-	-	0.0
96	67 years	NNN	-	-	667	-	-	-	0	-	-	0.0
98	21 years	NNN	-	-	1335	-	-	-	0	-	-	0.0
100	35 years	NNN	-	-	676	-	-	-	0	-	-	0.0
102	16 years	NNN	-	-	843	-	-	-	0	-	-	0.0
104	27 years	NNN	-	-	1003	-	-	-	0	-	-	0.0
106	65 years	NNN	-	-	1097	-	-	-	0	-	-	0.0
	35.0 years			(55)	(47662)	81.6%	5.7%	(207)	(24)	17.4%	49.2%	1.2

Table 8.8 Performance of the **Hybrid I** spike detection system (mimetic and rule based) on the **test data set**.

Recording				Stage 1 (Mimetic)				Stage 2 (Mimetic+Rules)				
EEG	Age	Type	Def	Mis	Fal	Sen	Sel	Mis	Fal	Sen	Sel	FD/hr
1	46 years	GGG	50	7	723	86%	6%	20	0	60%	100%	0.0
3	17 years	GGG	2	0	623	100%	0%	1	0	50%	100%	0.0
5	50 years	GGN	1	0	1101	100%	0%	1	0	0%	0%	0.0
7	56 years	GGN	3	0	922	100%	0%	3	1	0%	0%	2.5
9	2 weeks	MMM	56	41	226	27%	6%	54	6	4%	25%	22.4
11	65 years	GGL	15	2	744	87%	2%	14	0	7%	100%	0.0
13	7 months	GLF	-	-	778	-	-	-	0	-	-	0.0
15	33 years	NGG	-	-	327	-	-	-	0	-	-	0.0
17	25 years	MFN	-	-	368	-	-	-	1	-	-	2.3
19	2 years	NFN	-	-	238	-	-	-	0	-	-	0.0
21	2 years	NGN	-	-	23	-	-	-	0	-	-	0.0
23	62 years	GNN	-	-	1248	-	-	-	1	-	-	2.4
25	68 years	FNN	-	-	1606	-	-	-	3	-	-	6.8
27	70 years	NNF	-	-	1075	-	-	-	1	-	-	2.4
29	30 years	NNN	-	-	1479	-	-	-	1	-	-	2.5
31	48 years	NNN	-	-	898	-	-	-	1	-	-	3.3
33	10 years	NNN	-	-	1351	-	-	-	1	-	-	2.5
35	54 years	NNN	-	-	769	-	-	-	1	-	-	2.1
37	61 years	NNN	-	-	282	-	-	-	1	-	-	2.4
39	26 years	NNN	-	-	1106	-	-	-	1	-	-	2.3
41	6 years	NNN	-	-	1593	-	-	-	1	-	-	2.6
43	26 years	NNN	-	-	371	-	-	-	1	-	-	2.4
45	63 years	NNN	-	-	546	-	-	-	1	-	-	2.4
47	14 months	NNN	-	-	655	-	-	-	1	-	-	3.9
49	13 years	NNN	-	-	1006	-	-	-	2	-	-	5.6
51	7 years	NNN	-	-	1067	-	-	-	3	-	-	7.1
53	61 years	NNN	-	-	400	-	-	-	0	-	-	0.0
55	13 years	NNN	-	-	694	-	-	-	0	-	-	0.0
57	79 years	NNN	-	-	954	-	-	-	0	-	-	0.0
59	71 years	NNN	-	-	378	-	-	-	0	-	-	0.0
61	16 years	NNN	-	-	485	-	-	-	0	-	-	0.0
63	58 years	NNN	-	-	790	-	-	-	0	-	-	0.0
65	80 years	NNN	-	-	957	-	-	-	0	-	-	0.0
67	74 years	NNN	-	-	850	-	-	-	0	-	-	0.0
69	16 years	NNN	-	-	563	-	-	-	0	-	-	0.0
71	8 years	NNN	-	-	798	-	-	-	0	-	-	0.0
73	13 years	NNN	-	-	427	-	-	-	0	-	-	0.0
75	55 years	NNN	-	-	1151	-	-	-	0	-	-	0.0
77	28 years	NNN	-	-	1579	-	-	-	0	-	-	0.0
79	28 years	NNN	-	-	1019	-	-	-	0	-	-	0.0
81	5 years	NNN	-	-	682	-	-	-	0	-	-	0.0
83	39 years	NNN	-	-	1078	-	-	-	0	-	-	0.0
85	78 years	NNN	-	-	837	-	-	-	0	-	-	0.0
87	65 years	NNN	-	-	692	-	-	-	0	-	-	0.0
89	54 years	NNN	-	-	1069	-	-	-	0	-	-	0.0
91	5 years	NNN	-	-	450	-	-	-	0	-	-	0.0
93	6 weeks	NNN	-	-	83	-	-	-	0	-	-	0.0
95	23 years	NNN	-	-	908	-	-	-	0	-	-	0.0
97	12 months	NNN	-	-	672	-	-	-	0	-	-	0.0
99	66 years	NNN	-	-	1106	-	-	-	0	-	-	0.0
101	3 weeks	NNN	-	-	995	-	-	-	0	-	-	0.0
103	15 years	NNN	-	-	732	-	-	-	0	-	-	0.0
105	64 years	NNN	-	-	1343	-	-	-	0	-	-	0.0
	35.0 years			(50)	(42817)	83.2%	2.4%	(93)	(28)	20.0%	54.2%	1.5

Table 8.9 Performance of the **Hybrid II** spike detection system (mimetic and SOFM based) on the **training data set.**

Recording				Stage 3 (Mimetic+SOFM+Fuzzy)				
EEG	Age	Type	Def	Mis	Fal	Sen	Sel	FD/hr
2	12 years	GGG	22	12	0	45%	100%	0.0
4	16 years	GGG	3	2	0	33%	100%	0.0
6	8 years	MFM	189	189	0	0%	0%	0.0
8	84 years	MNF	4	4	0	0%	0%	0.0
10	5 years	GGG	5	1	1	80%	80%	2.2
12	4 years	MMM	36	36	0	0%	0%	0.0
14	37 years	NFF	-	-	4	-	-	9.8
16	25 years	GGN	-	-	-	-	-	-
18	49 years	NFN	-	-	23	-	-	51.7
20	84 years	NFN	-	-	-	-	-	-
22	3 years	NFN	-	-	8	-	-	30.8
24	29 years	FNN	-	-	1	-	-	2.6
26	20 years	NGN	-	-	26	-	-	57.2
28	57 years	NNN	-	-	43	-	-	102.4
30	42 years	NNN	-	-	64	-	-	149.8
32	6 months	NNN	-	-	0	-	-	0.0
34	29 years	NNN	-	-	6	-	-	14.5
36	25 years	NNN	-	-	-	-	-	-
38	35 years	NNN	-	-	65	-	-	177.8
40	7 months	NNN	-	-	1	-	-	3.4
42	43 years	NNN	-	-	0	-	-	0.0
44	14 years	NNN	-	-	0	-	-	0.0
46	14 years	NNN	-	-	0	-	-	0.0
48	65 years	NNN	-	-	65	-	-	137.4
50	25 years	NNN	-	-	27	-	-	64.0
52	32 years	NNN	-	-	20	-	-	43.8
54	62 years	NNN	-	-	0	-	-	0.0
56	32 years	NNN	-	-	1	-	-	2.3
58	65 years	NNN	-	-	4	-	-	9.6
60	26 years	NNN	-	-	16	-	-	38.4
62	14 years	NNN	-	-	0	-	-	0.0
64	73 years	NNN	-	-	0	-	-	0.0
66	13 years	NNN	-	-	0	-	-	0.0
68	8 weeks	NNN	-	-	1	-	-	4.4
70	30 years	NNN	-	-	45	-	-	107.3
72	62 years	NNN	-	-	0	-	-	0.0
74	77 years	NNN	-	-	0	-	-	0.0
76	43 years	NNN	-	-	1	-	-	2.4
78	62 years	NNN	-	-	0	-	-	0.0
80	48 years	NNN	-	-	0	-	-	0.0
82	73 years	NNN	-	-	0	-	-	0.0
84	48 years	NNN	-	-	14	-	-	33.7
86	8 years	NNN	-	-	0	-	-	0.0
88	40 years	NNN	-	-	2	-	-	4.6
90	20 years	NNN	-	-	12	-	-	28.1
92	60 years	NNN	-	-	2	-	-	4.8
94	12 years	NNN	-	-	1	-	-	2.3
96	67 years	NNN	-	-	0	-	-	0.0
98	21 years	NNN	-	-	0	-	-	0.0
100	35 years	NNN	-	-	27	-	-	64.0
102	16 years	NNN	-	-	3	-	-	7.1
104	27 years	NNN	-	-	1	-	-	2.3
106	65 years	NNN	-	-	0	-	-	0.0
	35.0 years			(244)	(484)	26.5%	46.7%	23.2

Table 8.10 Performance of the **Hybrid II** spike detection system (mimetic and SOFM based) on the **test data set**.

Recording				Stage 3 (Mimetic+SOFM+Fuzzy)				
EEG	Age	Type	Def	Mis	Fal	Sen	Sel	FD/hr
1	46 years	GGG	50	22	0	56%	100%	0.0
3	17 years	GGG	2	0	50	100%	4%	116.8
5	50 years	GGN	1	1	0	0%	0%	0.0
7	56 years	GGN	3	3	21	0%	0%	52.6
9	2 weeks	MMM	56	56	0	0%	0%	0.0
11	65 years	GGL	15	14	6	7%	14%	15.4
13	7 months	GLF	-	-	0	-	-	0.0
15	33 years	NGG	-	-	-	-	-	-
17	25 years	MFN	-	-	24	-	-	55.4
19	2 years	NFN	-	-	0	-	-	0.0
21	2 years	NGN	-	-	0	-	-	0.0
23	62 years	GNN	-	-	53	-	-	129.6
25	68 years	FNN	-	-	50	-	-	113.9
27	70 years	NNF	-	-	11	-	-	26.8
29	30 years	NNN	-	-	0	-	-	0.0
31	48 years	NNN	-	-	1	-	-	3.3
33	10 years	NNN	-	-	1	-	-	2.5
35	54 years	NNN	-	-	21	-	-	44.9
37	61 years	NNN	-	-	1	-	-	2.4
39	26 years	NNN	-	-	1	-	-	2.3
41	6 years	NNN	-	-	5	-	-	12.9
43	26 years	NNN	-	-	5	-	-	12.0
45	63 years	NNN	-	-	1	-	-	2.4
47	14 months	NNN	-	-	0	-	-	0.0
49	13 years	NNN	-	-	0	-	-	0.0
51	7 years	NNN	-	-	1	-	-	2.4
53	61 years	NNN	-	-	0	-	-	0.0
55	13 years	NNN	-	-	70	-	-	163.1
57	79 years	NNN	-	-	0	-	-	0.0
59	71 years	NNN	-	-	0	-	-	0.0
61	16 years	NNN	-	-	-	-	-	-
63	58 years	NNN	-	-	1	-	-	2.3
65	80 years	NNN	-	-	0	-	-	0.0
67	74 years	NNN	-	-	0	-	-	0.0
69	16 years	NNN	-	-	0	-	-	0.0
71	8 years	NNN	-	-	0	-	-	0.0
73	13 years	NNN	-	-	3	-	-	7.1
75	55 years	NNN	-	-	0	-	-	0.0
77	28 years	NNN	-	-	4	-	-	9.6
79	28 years	NNN	-	-	1	-	-	2.4
81	5 years	NNN	-	-	0	-	-	0.0
83	39 years	NNN	-	-	0	-	-	0.0
85	78 years	NNN	-	-	14	-	-	32.3
87	65 years	NNN	-	-	114	-	-	276.4
89	54 years	NNN	-	-	4	-	-	9.3
91	5 years	NNN	-	-	0	-	-	0.0
93	6 weeks	NNN	-	-	0	-	-	0.0
95	23 years	NNN	-	-	2	-	-	4.7
97	12 months	NNN	-	-	3	-	-	11.1
99	66 years	NNN	-	-	0	-	-	0.0
101	3 weeks	NNN	-	-	0	-	-	0.0
103	15 years	NNN	-	-	0	-	-	0.0
105	64 years	NNN	-	-	0	-	-	0.0
	35.0 years			(96)	(468)	27.1%	19.7%	21.8

of 27.1%, a mean selectivity of 19.7%, and a false detection rate of 21.8 per hour on the current test set (Table 8.10).

The Hybrid II system had previously been evaluated with 8 recordings (7 epileptiform, 1 normal) in a blind study yielding a mean sensitivity of 55.3%, a mean selectivity of 82.0%, and a false detection rate of 7.2 per hour [James *et al.* 1999]. The recordings used had been evaluated by 2 to 3 readers. It is unclear why there is such a striking difference in performance of Hybrid II between the previous test data set (8 EEGs) and the current test data set (53 EEGs) but, at least in part, this may be attributed to the use of a different version of the Hybrid II system, differently trained SOFMs, and substantially different test sets.

The results emphasize the importance of a large set of normal controls as the variety of normal EEG patterns which may potentially be misclassified by the system is quite large. It appears that only if the patterns are part of the training set can they be rejected by the system successfully.

8.7 PERFORMANCE AT THE LEVEL OF ENTIRE RECORDINGS

Performances of the three spike detection systems at both individual-spike and global-EEG levels with respect to the entire data set of 106 EEGs with 27 epileptiform EEGs and 79 normal controls are summarized in Table 8.11. Mean sensitivities and selectivities[4] are listed with respect to the 12 EEGs containing definite EDs (those marked as definite by at least two of the readers). An EEG was considered definitely epileptiform if it was marked as such by two or three readers (17 out of 106). If only one reader reported epileptiform activity, the EEG was neither considered epileptiform nor normal (10 out of 106). EEGs labelled non-epileptiform by all 3 EEGers were considered normal (79 out of 106).

For the Hybrid I and II systems three separate evaluations were made in which 'possible', 'probable' and 'definite' events were counted as detections. Stage one detections are included for the Hybrid I system and the wavelet-based system. Only 101 EEGs were successfully evaluated with the Hybrid II system. Definitely epileptiform EEGs without detection were counted as global missed detections ('Mis EEG' in Table 8.11). Normal EEGs with one or more repeated detections were counted as global false detections ('Fal EEG' in Table 8.11).

The Hybrid I system features a very low false detection rates (0.2 to 1.3 FD/h). In the most sensitive mode (counting possible EDs) it misses four of the epileptiform

[4]Since sensitivities and selectivities of individual EEGs were averaged, a seemingly paradoxical situation of decreasing sensitivity *and* selectivity arises for the possible and probable stage-2 detections of the Hybrid I system in Table 8.11.

Table 8.11 Mean performance of three spike detection systems: Hybrid I, Hybrid II and the wavelet-based system. Listed are the sort of detection (Det), the number of definite EDs correctly detected (Cor), the number of definite EDs missed (Mis), the number of false detections (Fal), the mean sensitivity (Sen), the mean selectivity (Sel), the mean hourly false detection rate (FD/hr), the number of EEGs processed (EEGs), the number of definitely epileptiform EEGs without detections (Mis EEGs), and the number of definitely normal EEGs with detections (Fal EEGs). The sort of detection used for the compilation of the previous tables is marked by (*).

System	Det	Individual Events						Entire EEGs		
		Cor	Mis	Fal	Sen	Sel	FD /hr	EEGs	Mis EEGs	Fal EEGs
Hybrid I	Stage 1	281	105	90479	82.4%	4.0%	2155.4	106	-	-
	Stage 2 Poss (*)	86	300	52	18.7%	88.6%	1.3	106	4	25
	Stage 2 Prob	74	312	20	15.8%	87.3%	0.5	106	8	6
	Stage 2 Def	32	354	8	13.6%	100%	0.2	106	9	5
Hybrid II	Stage 3 Poss	101	285	39128	45.5%	29.4%	919.1	101	1	71
	Stage 3 Prob (*)	46	340	952	26.8%	66.4%	22.5	101	5	41
	Stage 3 Def	8	378	0	5.8%	100%	0.0	101	12	0
Wavelet based	Stage 1	323	63	47587	95.7%	15.1%	1178.4	106	-	-
	Stage 2 Def (*)	131	255	258	57.9%	62.5%	6.8	106	1	40

EEGs (8, 14, 15, and 16). In the most selective mode (counting definite EDs only) there are five false detections (EEGs 29, 30, 31, 32, and 33).

The Hybrid II system misses one recording (EEG 9, p. 175) in the most sensitive mode but falsely marks 71 out of 79 normal EEGs. In the most selective mode it operates without false detections but misses 12 out of 17 definitely epileptiform EEGs.

The wavelet-based system features the highest mean sensitivity of 57.9% and does so with a reasonably high selectivity of 62.5%. It missed 1 of the epileptiform EEGs (EEG 16, Fig. A.9 on p. 164) and reported more than half of the normal EEGs as epileptiform.

All three systems took only 5-10% of the recording time for the evaluation on a Pentium II running at 500 MHz. Thus all systems could potentially operate in real-time mode. A real-time implementation of the Hybrid I system has been developed [Green 1995, Jones *et al.* 1996].

For the wavelet-based system thresholds for primary detection (0.66 in Fig. 7.6), final classification (Eq. 7.35) and focal upgrade (Eq. 7.36) were chosen to balance sensitivity and selectivity. These thresholds can be adjusted to optimize either the sensitivity or the selectivity of the the system. Thus, the system can be adapted to work as a screening tool or a long-term monitoring system as outlined in section 2.5.

8.8 PERFORMANCE OF SYSTEMS IN THE LITERATURE

Pietilä *et al.* [1994] compared two systems developed by Värri *et al.* [1988] and Gotman and Gloor [1976] with 12 recordings from 6 epileptic patients. For each patient one recording was made while they were awake and one sleep recordings was used. They report a sensitivity of 31% and a specificity of 17% for the Tampere (Finland) based system [Värri *et al.* 1988]. On the same dataset the system developed by Gotman [1985] achieved a sensitivity of 33% and a specificity of 49%.

Tong *et al.* [1996] tested their wavelet-based spike detection system on 61 min of EEG (recording length: 4 min) taken from 17 patients. They reported successful detection of 177 of 187 EDs (sensitivity: 95%) while only 22 false detections were reported (selectivity: 89%, false detection rate: 22 per hour).

Sartoretto and Ermani [1999] used 79 min of EEG with 48 EDs from 8 patients to evaluate their wavelet-based system. They reported a sensitivity of 96%, a selectivity of 37% and a false detection rate of 93 per hour. Their system is based on the DWT which makes it considerably faster than the one presented in this work; for example, analysis of a 310 s 8-channel-recording took only 1.0 s.

A two-stage wavelet-based system developed by Senhadji *et al.* [1995] had a sensitivity of 86% and selectivity of 93% over 982 EDs [Senhadji *et al.* 1994].

Dümpelmann and Elger [1999] analyzed 136 min of intracranial EEG from seven patients under presurgical evaluation. EEGs were rated by two readers and three spike detection systems: a system developed by Gotman and Gloor [1976], a two-stage system based on an autoregressive model [da Silva *et al.* 1975] and a wavelet-based system similar to the one proposed by Kalayci and Özdamar [1995]. They reported sensitivities of 23.0-25.1% for the Gotman system, 31.5-33.5% for the autoregressive system and 24.8-27.8% for the wavelet-based system.

Ramabhadran *et al.* [1999] tested a multistage spike detection system with mimetic feature extraction and rule-based classification with 13 epileptiform recordings and 5 normal controls that included all stages of sleep and wakefulness. They reported a sensitivity of 95.7 % and a selectivity of 88.9%. All false detections were eliminated by rule-based evaluation of the wide-sense temporal context of CED foci. Thus, no false detections occurred within 3 hours of normal control recordings.

The number of patients has not been reported for all systems in the literature. However, the importance of using a large set of normal controls with a variety of artifacts to identify potential 'leaks' in the discrimination process has been demonstrated in the evaluation of three systems. All age groups were represented in the data set used here. Specifically, EEGs of young children may contain movement and electrode artifacts that are difficult to discriminate from EDs. It is questionable whether other systems reported in the literature would achieve similar performances on the data set used in this work.

8.9 POSSIBLE IMPROVEMENTS

The wavelet-based spike detection system presented in this work would benefit from further refinement before being commissioned for clinical applications. The system was defined in a straightforward manner as a nonlinear multichannel filter. Features were selected and adjusted in the design and training process to optimize the ANOVA F-ratios of the discriminant functions with respect to the training data set. The feature-specific F-ratios were optimized by adjusting

- the parameter α for the spatial analysis, and
- the sequence of the chirp filter.

After each change, an evaluation of the entire training set was required. Depending on the resulting F-ratio, the change was subsequently implemented or rejected.

The discriminative quality of features can be illustrated by their F-ratios. The F-ratio of the linear discriminant rises if further features are added. However, whether or not a specific new feature makes a novel contribution to the discriminant function is measured by its F-remove value - the reduction of the F-ratio of the discriminant function caused by the removal of the feature. F-ratios and F-remove values are listed in tables 8.12 and 8.13 for single-channel features and multichannel features, respectively. The chirp features are the best discriminating single-channel features.

Table 8.12 F-ratios of single-channel features and F-remove values with respect to the linear discriminant function $g(X)$.

Single-Channel Features			F=3996
Feature	Symbol	F-ratio	F-remove
ridge count scales 6,8	$s_{i(1)}$	43.1	19.8
ridge count scales 12,16	$s_{i(3)}$	455.8	200.4
ridge count scales 24,32	$s_{i(5)}$	437.9	58.8
ridge count scales 48,64	$s_{i(7)}$	532.5	303.3
ridge count scales 96,128	$s_{i(9)}$	271.0	127.4
mean amplitude scales 24,32	$m_{i(5)}$	227.6	89.8
mean amplitude scales 48,64	$m_{i(7)}$	460.2	957.2
mean amplitude scales 96,128	$m_{i(9)}$	156.3	973.1
EOG	$f_i^{[\mathrm{EOG}]}$	63.8	42.5
chirp	$f_i^{[\mathrm{chirp1}]}$	1595.2	506.0
chirp ridge	$f_i^{[\mathrm{chirp2}]}$	850.4	173.7
adjacent dist	$f_i^{[\mathrm{dist}]}$	107.2	123.7
sharp	$f_i^{[\mathrm{sharp}]}$	73.8	386.7

However, in conjunction with the other features, the mean amplitude of scales 96 and 128 makes the largest contribution (closely followed by the mean amplitude of scales 48 and 64). The system would suffer the smallest loss if the ridge count of scales 6 and 8 was removed. The RMS probabilities is the best discriminating multichannel feature. It may be possible to find a feature that identifies the artifacts that cause

Table 8.13 F-ratios of multichannel features and F-remove values with respect to the linear discriminant function $g'(X')$.

Multichannel Features			F=4958
Feature	Symbol	F-ratio	F-remove
RMS probabilities	$f^{[\mathrm{prms}]}$	4031.0	2508.5
2nd largest probability	$f^{[\mathrm{p2}]}$	1148.8	237.0
source orientation estimate	$f^{[\mathrm{rad}]}$	234.1	240.9
adjacency filter	$f^{[\mathrm{adj}]}$	1970.1	297.2

large numbers of false detections, e.g., in recordings 22, 40 , 42 and 70. If the new feature was added to the system, one of the present features may become redundant and could be removed.

For clinical use in a long-term monitoring system, the training set would have to be augmented by representative samples of sleep recordings, and additional features may be needed for their discrimination from EDs. A state-dependent classification approach might prove beneficial [Gotman and Wang 1991, Gotman and Wang 1992].

Importantly, an additional stage evaluating the wide-sense temporal context should be added. The global wide-sense temporal context could help considerably in the evaluation of focal events. Epileptic foci can only be verified by a statistical evaluation of the spatial distribution of detected EDs. Each detection would be associated with the most probable location of a tentative focus. The location could be referenced by either a channel, an electrode, or a node of the geodesic grid. A linear classification function could be used to declare a focus, based on the properties of focal and non-focal EDs in the training set. Once a focus is declared in a recording, the Stage 1 detections (or, possibly, the entire recording) would be re-assessed and upgraded if compatible with the spatial characteristics of the focus.

The data set contained several epileptiform bursts and *seizures* (e.g. Fig. A.1 on page 156) all of which featured spikes and thus were detected by the system. On average, a number of 3-5 spikes were detected at the onset of a seizure. The detection threshold in the single-channel analysis (Fig. 7.5) is derived from the floating mean of the preceeding 5 s. If seizures last for several seconds, this threshold rises lowering the system's sensitivity. Thus spikes in the later part of a seizure may not be detected by Stage 1.

It would be advantageous to add the subjects' age group (e.g., newborn, under 5, and 5 and above) to the prior knowledge of the system since it has a large influence on the characteristics of normal EEG patterns. Separate classifiers could be trained and applied within each age group for the wavelet-based system. However, for each age group a substantial data set would be required.

Chapter 9

CONCLUSIONS AND FUTURE RESEARCH

9.1 CONCLUSIONS

In wavelet-based signal analysis there are two different basic approaches: DWT and CWT. The DWT is implemented with a fast algorithm that repeatedly splits the spectrum into an approximation signal and a detail signal. The approximation signal is downsampled at each level. The transform can be inverted and perfectly reconstructs the original signal. However, the DWT is not translation-invariant which is disadvantageous for some detection applications since the way a transient is represented depends on its temporal alignment with respect to the DWT.

The CWT is a redundant transform that can be continuous in both time and scale. It operates similar to a series of bandpass filters and is translation-invariant since no downsampling occurs. A fixed prototype filter, the analyzing wavelet or wavelet filter, is re-scaled to cover a predefined bandwidth. The wavelet filter must be chosen for a specific application and can be real-valued or complex-valued. A crucial feature of a wavelet filter is good time-frequency localization. Thus, most common choices are derivatives of the Gaussian and modulated Gaussians, including both real- and complex modulation. Modulated window functions other than the Gaussian can provide a smaller relative bandwidth, which is crucial for the analysis of transient signals. A very small relative bandwidth can be achieved with a complex modulated sine halfwave and a complex modulated Hanning window. The complex-valued wavelet makes the CWT independent of the phase and thus allows for the analysis of the time-frequency content of a signal. However, the traditional bilinear time-frequency distributions provide superior time-frequency resolution.

Features of transients can be extracted from the temporal local maxima of the CWT or multiscale edges. The amplitude characteristic of a multiscale edge can provide information about the frequency content of a very short transient such as a single half-wave or an edge. In contrast, the local maxima of the CWT with respect to scale, or multiscale ridges, provide useful indicators of the instantaneous frequency of brief oscillations.

The field of statistical pattern recognition includes a multitude of classification methods such as fuzzy logic and artificial neural networks. Classic linear and non-linear discriminant functions perform supervised learning from a labelled training set. The output of discriminant functions can be improved with a Bayesian approach to obtain a measure of certainty for the class membership of a feature vector, similar to the defuzzification step in fuzzy logic.

In the spatial analysis of the EEG, three basic spaces need to be considered: source space, electrode space and channel space. Dipolar current sources in an electrostatic model of the head provide a model for the generation of the electric surface potential of the EEG. The potential field, as sampled by a number of electrodes, is mapped into channel space by one of several montages. This transformation from electrode space to channel space cannot be inverted due to the lack of an absolute zero reference. Depending on the choice of montage, the EEG becomes sensitive to specific regions of the source volume. The hemispheric source volume can be sampled with a geodesic grid derived from a set of concentric n-frequency icosahedra. The more accurately a putative source at any grid location can account for the surface potential, the more likely it represents a current generator. Location and orientation of an unknown source can be estimated from the goodness-of-fit obtained for all grid nodes. The estimated source orientation with respect to spherical coordinates is a valuable feature for the discrimination of EDs from artifacts.

For the background EEG, a chi-square distribution can be fitted to the magnitude of the corresponding wavelet coefficients on a logarithmic scale. Transients such as interictal EDs can be detected as deviations of the log wavelet coefficients from the mean of the preceding 5 s at a scale centered around 17 Hz. For each detection, a set of features extracted from the raw signals and the CWTs of all channels provides a means of discriminating EDs from artifacts. Features extracted from ridges of the CWT help discriminate ETs from oscillatory activity such as alpha waves. Two features relating to a diagonal or chirp-like amplitude distribution of ETs in the time-scale plane (scale increasing with time) of the CWT are characteristic for ETs with respect to other spike-like transients. The sharpness of a transient, defined by the amplitude of the corresponding multiscale edge at the smallest scale, helps reject sharp artifacts such as muscle spikes and electrode-'pops'. Low frequency components of slow eye-blinks or eye-movements are captured by the wavelet coefficient of the scale centered around 1.6 Hz. For the final discrimination of an EDs from artifacts it is advantageous to evaluate the RMS spike probability of all channels, the source orientation estimate obtained with the spatial model and the adjacency pattern of observed polarities and single channel spike probabilities. The ANOVA F-ratio of a linear discriminant function with respect to a training data set is an important indicator for optimal feature selection and adjustment.

The performance of the wavelet-based spike detection system compares favourably with the Hybrid I and Hybrid II systems in terms of detection of definite EDs. However, in the evaluation of entire recordings (the global EEG level) the Hybrid I systems makes use of the wide-sense temporal context leading to a small number of normal EEGs being erroneously classified as epileptiform. None of the systems succeeds in finding all of the 17 definitely epileptiform recordings in the data set.

9.2 SUGGESTIONS AND FUTURE RESEARCH

9.2.1 Wavelet Analysis in the Detection of Epileptiform Spikes

Wavelet-based methods in general, and the CWT in particular, have a very close relationship to both human physiology and quantum physics. For example, Mallat [1989b] showed that wavelets can approximate the early stages of the human visual and auditory systems. At the same time, the CWT is part of the affine class of time-frequency distributions which originated in quantum physics [Quian and Chen 1996]. This wide-ranging background makes the CWT an attractive tool in detection and classification problems such as the detection of epileptiform spikes. Frisch and Messer [1992] also found that the CWT can provide an optimal detector for a largely unknown transient.

The DWT provides a different approach that involves downsampling of the filtered signals. The DWT has been applied to the spike detection problem by a number of authors [Kalayci *et al.* 1994, Kalayci and Özdamar 1995, D'Attelis *et al.* 1997, Kim *et al.* 1998, Sartoretto and Ermani 1999]. It was not used in the work reported here because of its lack of translation invariance which can have dramatic effects on the output of the transform as shown in Fig. 4.6. Oversampling methods for the DWT as proposed by Sari-Sarraf and Brzakovic [1997] have not been applied to the spike detection problem yet. The Daubechies wavelets may be useful because they show a chirp feature as indicated by the rising frequency in Fig. 3.4 ($n = 28$). Since the time-inverse of the rising frequency feature (increasing scale) has been found to be a characteristic feature of most ETs in this study. Thus, a matched filter could be constructed with the Daubechies-wavelet and an oversampled DWT.

The CWT has been applied to the spike detection problem by several authors [Schiff *et al.* 1994b, Senhadji *et al.* 1994, Senhadji *et al.* 1995, Clarencon *et al.* 1996, Tong *et al.* 1996, Popescu 1998, Goelz *et al.* 2000b]. Schiff *et al.* [1994b] used a real-valued wavelet to study spikes and seizures whereas all others have applied a complex-valued wavelet with the CWT. Only Senhadji *et al.* [1995] used a complex-valued wavelet with a small relative bandwidth matched to ETs to detect spikes. Complex-valued wavelets have been shown to optimally detect transients of unknown phase [Frisch and Messer 1992].

In the work reported here, the statistical distribution of the EEG's wavelet coefficients is modelled on a log scale. Among the statistical moments of the log distribution, only the mean relates to the amplitude of the EEG signal. All other statistical moments (such as variance, skewness, kurtosis, etc) provide information about the non-linear behaviour of the EEG. These features could potentially be used in the detection of seizures and may also provide clues for their prediction. Other non-linear features like the *Lyapunov exponent* and the *correlation dimension* have been applied to seizure prediction in cortical EEG recordings successfully [Iasemidis *et al.* 1996, Lehnertz and Elger 1998].

To measure the sharpness of transients the magnitude of the smallest scale wavelet coefficient of the corresponding multiscale edge was used in this work (Eq. 7.16). In order to obtain a measure for the sharpness of a transient that is independent from its amplitude a (linear or low-order polynomial) regression of the wavelet coefficients of the the multiscale edge could be used.

If a sufficiently large amount of computational power was available, matching pursuit (see subsection 4.4) could be applied to the spike detection problem. The waveform dictionary could be split into two parts representing EDs and artifacts. The overall variance covered by those atoms that correspond to EDs could be a valuable indicator for EDs.

9.2.2 Chirps

The chirp filters (see section on Chirp features in Chapter 7) based on the CWT ridges correspond to a characteristic feature of most ETs which could be called *chirp feature.*

In spike detection systems, wavelet transforms are used at the lowest level before feature extraction takes place. As the results for the Stage 1 detections in Table 8.11 (p. 145) indicate, the wavelet-based approach presented in this work can provide a very sensitive and selective filter for EDs compared to the mimetic approach which was designed to mimic the reader's perception. Furthermore, a feature like the chirp feature of ETs can only be observed using a time-frequency or time-scale representation such as the CWT.

Time-frequency signatures of seizures recorded with depth electrodes were presented by Senhadji *et al.* [1998], where several of the signatures showed decreasing frequency. Similarly, spectrograms of ictal cortical grid recordings by both Franaszczuk *et al.* [1998] and Schiff *et al.* [2000] show time-frequency components with decreasing frequency. Schiff *et al.* [2000] called these components 'brain chirps'.

While the frequencies of seizures decrease over periods of several seconds, the chirp feature of ETs relates to transients with only a few half waves. The CWT ridges of spike-wave complexes (Fig. 7.11, top centre) show a large overlap with ridges of alpha spindles (Fig. 7.11, bottom left). Thus, if a 'cortical resonant circuit' as proposed by Nunez [1981] exists, a spike-wave complex could be interpreted as the impulse response of such an intracortical network that would normally oscillate at the frequency of alpha waves. It would be interesting to investigate whether series of spike-wave complexes as found in epileptic seizures represent pathological oscillations of these putative networks.

The path of the chirp in the time-scale plane was modelled by the sequence defined in Eq. 7.20. The chirp features were found to provide the best discrimination between spikes and artifacts (Table 8.12). In order to further improve the performance of the chirp filters, a two-dimensional distribution based on a small number of parameters could be used to model the chirps. The parameters of the distribution could be adjusted to optimize the differentiation of ETs from artifacts.

9.2.3 Discriminant Functions

The quadratic discriminant function provides better classification results than the linear discriminant function as shown in the simulations in Chapter 5. However, it was not implemented in the system proposed in this work. To train a quadratic discriminant function, the covariance matrix for ET features needs to be estimated. Thus, for the 13-dimensional feature space of the single-channel discriminant function, 91 unique coefficients in the covariance matrix are sought. Since there were only 360 ETs in the training sample, the estimation could not be made reliably. If, however, a larger training set was available, the quadratic discriminant function could be implemented.

The training of a discriminant function is a straightforward calculation. In contrast, the training of an ANN is a lengthy iterative procedure. The training vectors have to be presented repeatedly to the ANN and adjustments have to be made during each iteration. However, an ANN can adapt to highly non-linear features such as varying scale in different regions of a training data set.

The Bayesian approach to labelling the output of the discriminant functions in the system presented leads to continuous-valued outputs which are compared to thresholds that trigger Stage 1 and Stage 2 detections (0.66 in Fig. 7.6 for Stage 1 detections, 0.7 in Eq. 7.35 for the final classification and 0.6 and 0.8 in Eq. 7.36 for a focal upgrade). These thresholds can be adjusted to optimize the system as a screening tool or a long-term monitoring system (see section 2.5). Two sets of thresholds could be build into the system such that the user could switch between both applications by turning a single 'knob'.

9.2.4 Knowledge

In the system presented in this work, the linear discriminant functions provide a means of classification of CEDs with additional confidence levels derived from a Bayesian approach. The detection threshold was chosen to make the system very sensitive but this was achieved at a price of a high false detection rate if certain artifacts coincided in a recording.

The wavelets can be considered to sense single-channel transients and the linear discriminant functions provides some form of inherent 'knowledge' about ETs relative to non-epileptiform transients.

In contrast, a rule-based expert system, as in Stage 2 of the Hybrid I system, is capable of analyzing the wide-sense temporal context of detected EDs. This context can be evaluated to eliminate false detections or verify epileptic foci [Davey *et al.* 1989, Glover *et al.* 1989, Dingle *et al.* 1993, Ramabhadran *et al.* 1999].

As mentioned in the previous chapter, any misclassification can be traced back in a rule-based system. The knowledge present in a rule-based system is explicitly readable while the 'knowledge' of statistical classifiers is represented implicitly by a set of weight vectors.

A rule-based expert system could be added to the wavelet-based system. The output of the single-channel or multichannel statistical classifiers would not provide sufficient information for meaningful rule-based evaluation. Thus, the part of the rule-base relating to single-channel analysis would have to be redeveloped to match the features extracted from the scalogram (Chapter 7). These rules could be considered low-level knowledge. Rules used in the analysis of the wide-sense temporal context could be added to the wavlet-based system as an additional stage (Stage 3).

Research in the field of expert systems is being conducted to find ways of extracting explicit knowledge from ANNs [Fu 1994]. It may be possible to develop a similar technique to extract rules (or even fuzzy rules) from a statistical classifier.

The rule-base of the Hybrid I system relating to the wide-sense temporal context could be adaptable to the wavelet-based system since these rules represent high-level knowledge.

Appendix A

RELEVANT EEG EPISODES

Charts of EEG episodes relevant to the evaluation of the spike detection systems in Chapter 8 are catalogued in a series of Figures on the following pages. The horizontal scale is fixed to 3cm/s with 5 s (half a page in the original paper recordings) shown in each figure. The vertical scale is either 100 or 200μV/cm, depending on the gain setting used in the recording. Each chart features a complete reference to the EEG number, the patient's age, the run and page number of the paper recording, the sample number of the digital recording and the current montage. Charts are sorted by increasing EEG numbers.

Detections of the wavelet-based spike detection system are superimposed in each chart if present. A dot at the baseline of each channel denotes the fiducal point and a number indicates the spike probability for each single-channel event. The multichannel probabilities of all multichannel events in a chart is listed in the top left corner (e.g. p=0.87).

E776C, Age: 17 years, Run 9, Page 378.95, Sample 190000, Montage: longitudinal

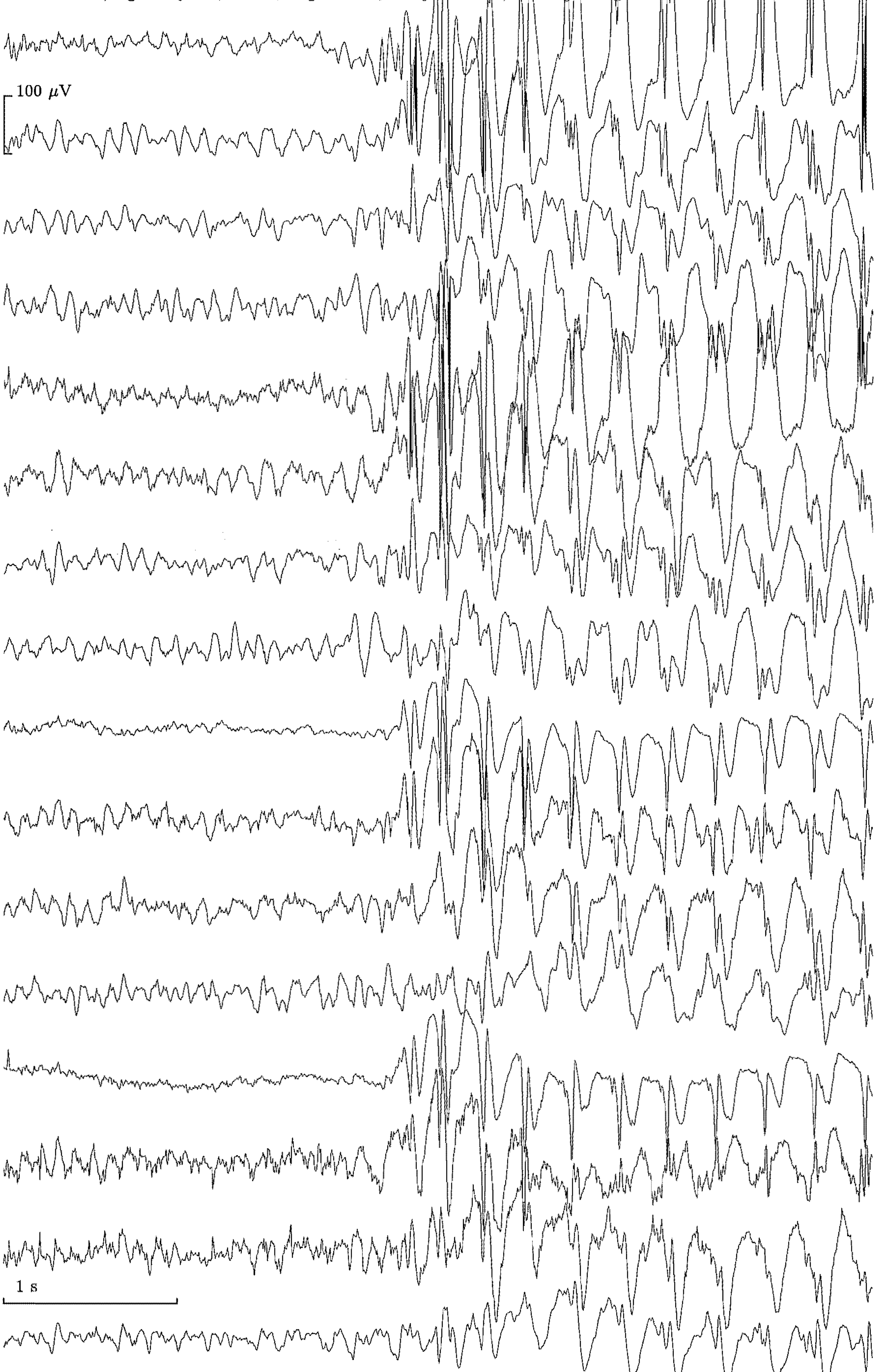

Figure A.1 Onset of a seizure. A series of spike and wave complexes at ca. 3.5 per second starts. Detections of the wavelet-based system are not shown (p. 12).

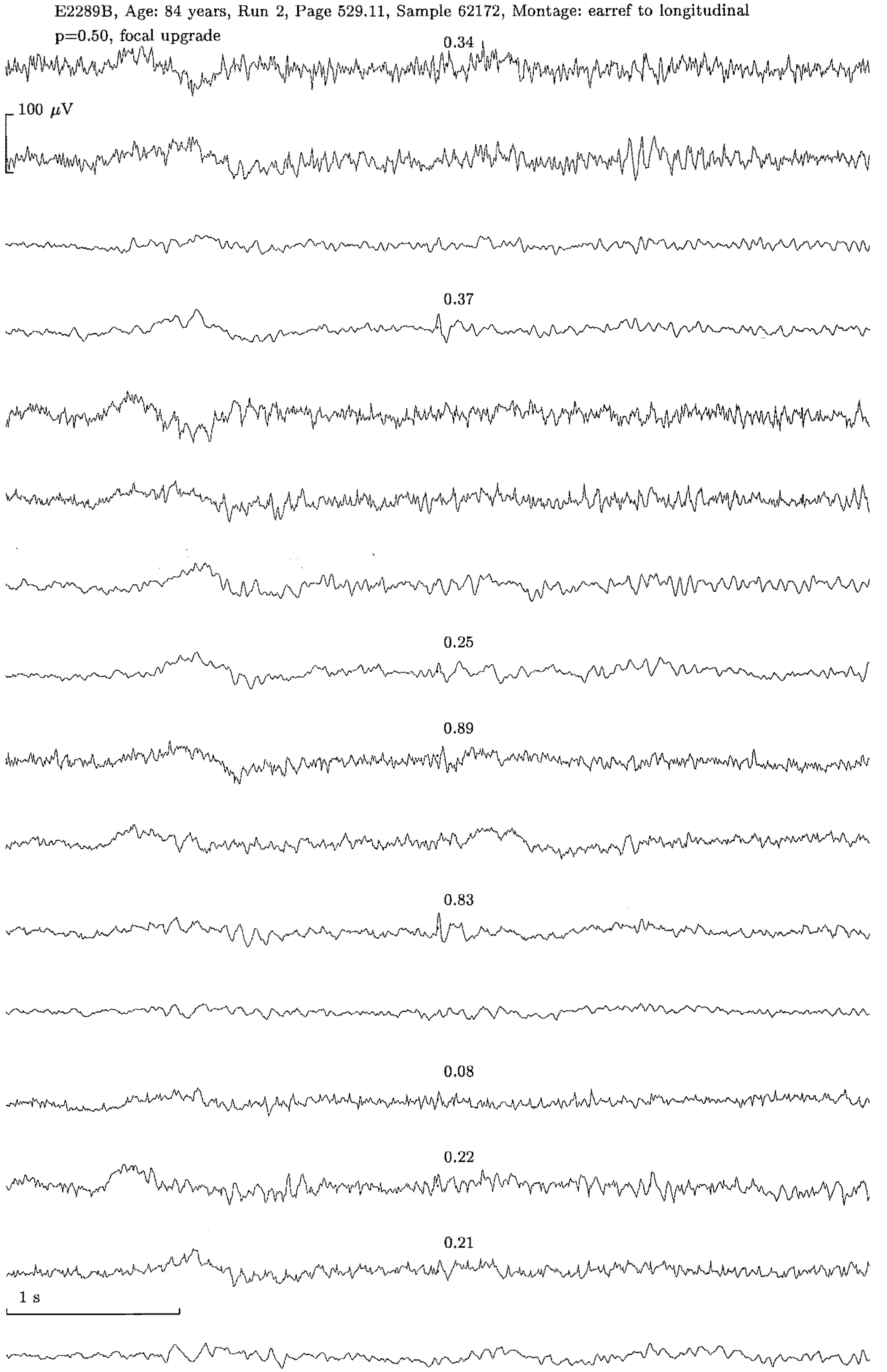

Figure A.2 A definite ED in EEG 8 (E2289B) was attenuated through the re-referencing but marked as definite by the wavelet-based system. Single channel probabilities on channels 9 and 11 were above 80% and, thus, the event qualified for upgrading as a focal event (p. 138).

E3394B, Age: 12 years, Run 1, Page 901.79, Sample 9000, Montage: longitudinal

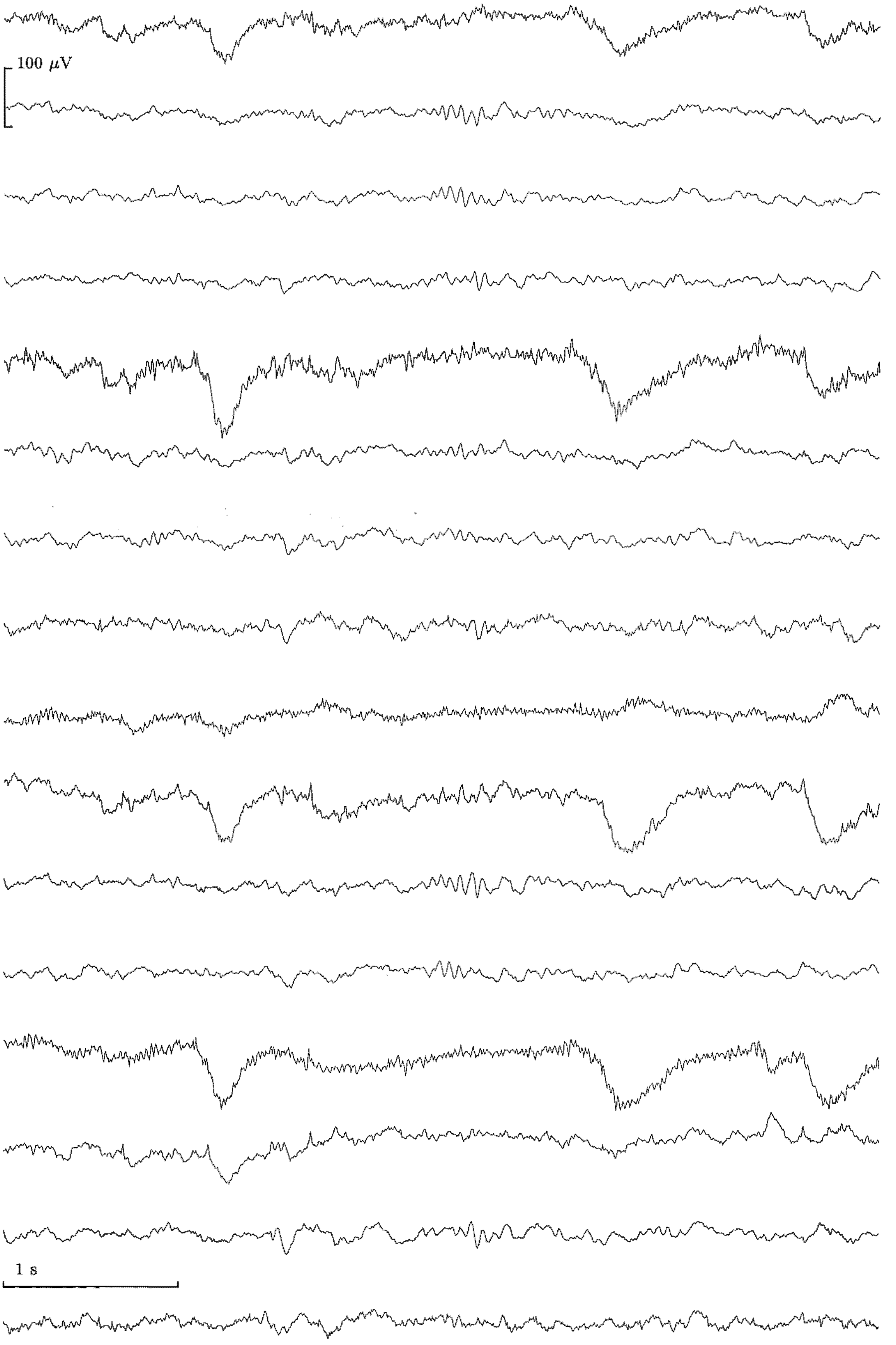

Figure A.3 Burst of beta activity (p. 8).

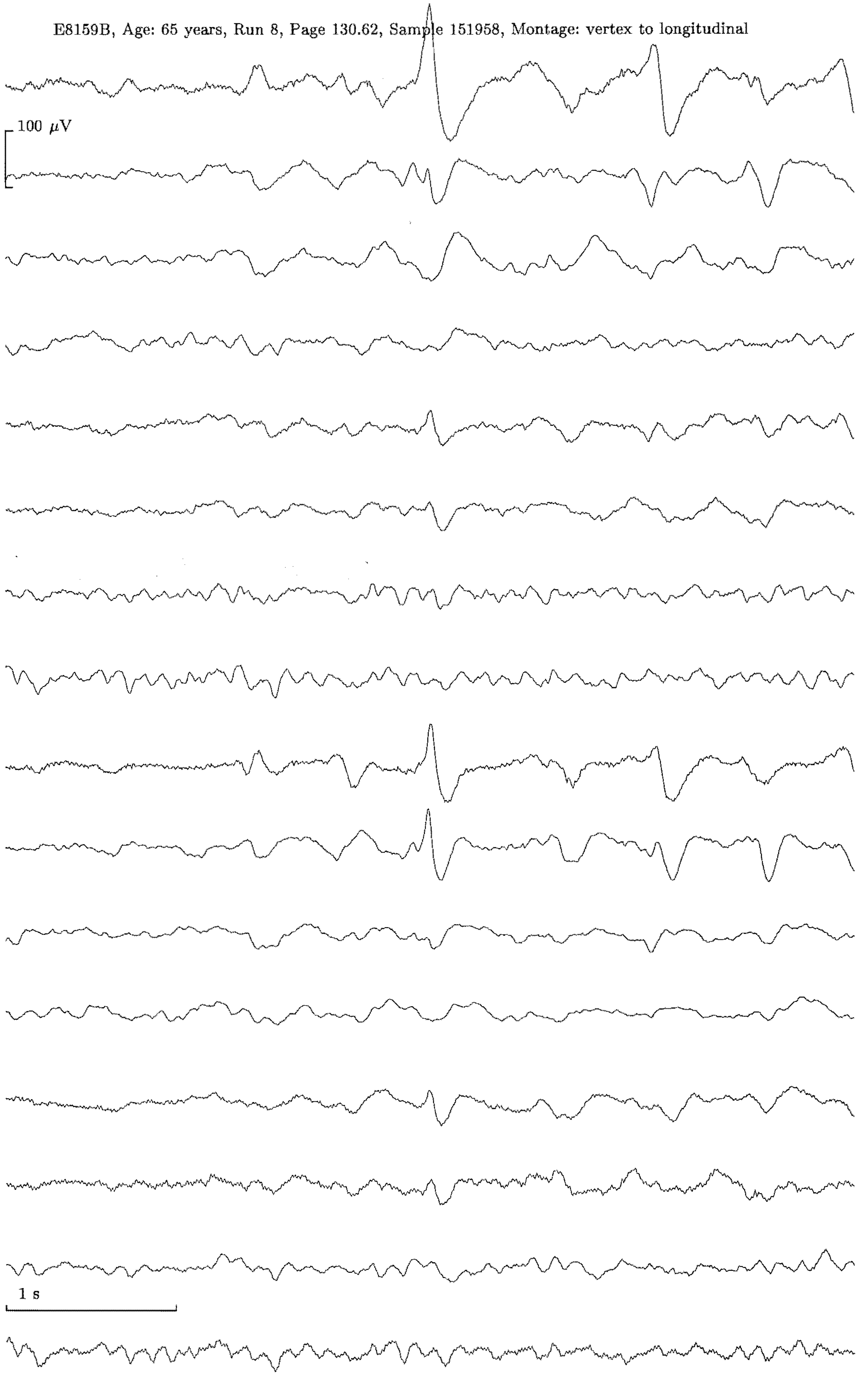

Figure A.4 This epileptiform sharp wave in EEG 11 (E8159B) was marked as questionable by the Hybrid I system. None of the 15 definite EDs in this recording were picked up by the wavelet-based system. Amplitude and scale of the sharp waves are similar to those found in fast eye-blinks (p. 174). Similar epileptiform activity did not occur in the training set (p. 135).

E8214, Age: 62 years, Run 7, Page 417.57, Sample 145352, Montage: earref to longitudinal

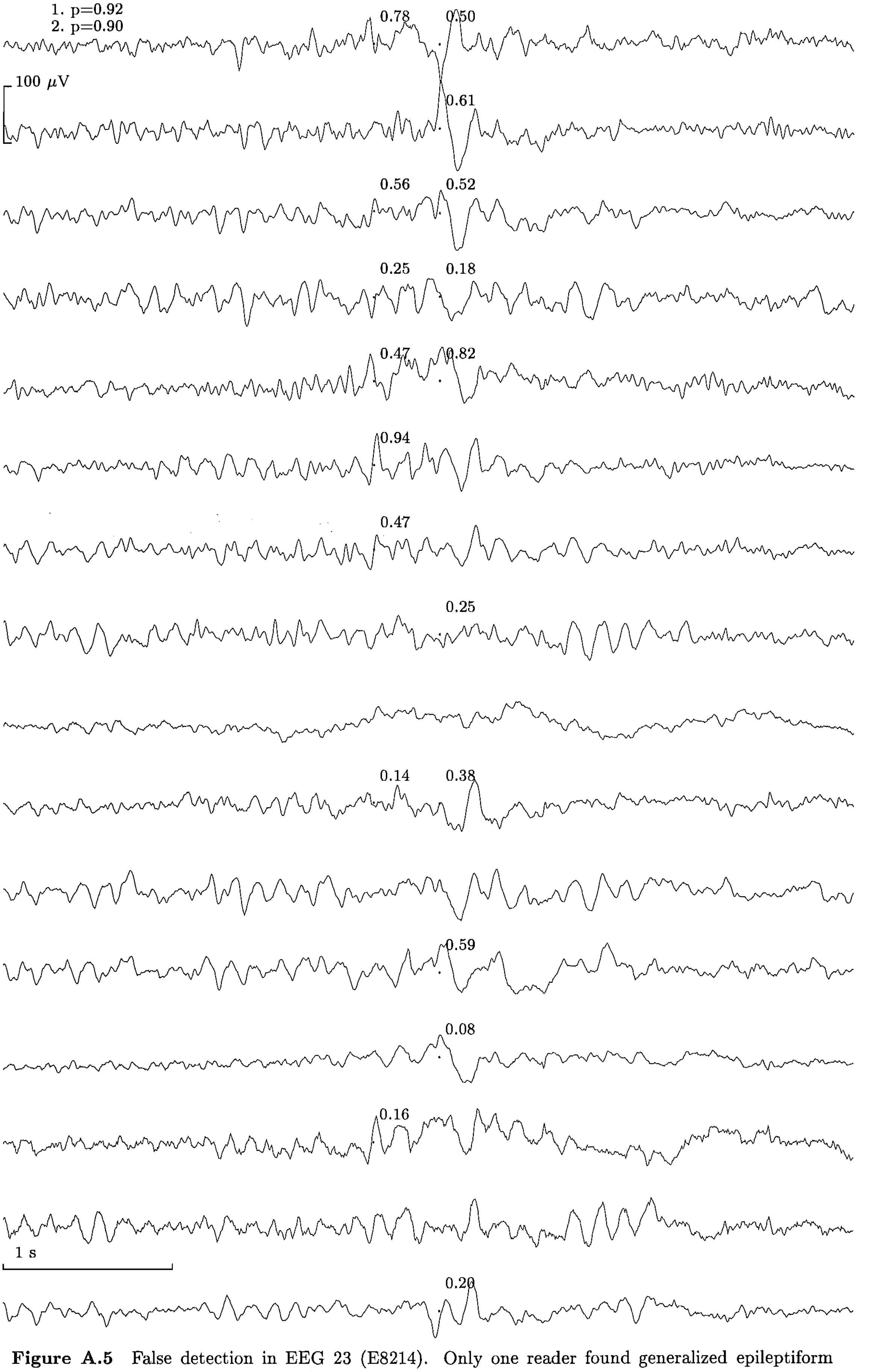

Figure A.5 False detection in EEG 23 (E8214). Only one reader found generalized epileptiform activity elsewhere in this recording (p. 138).

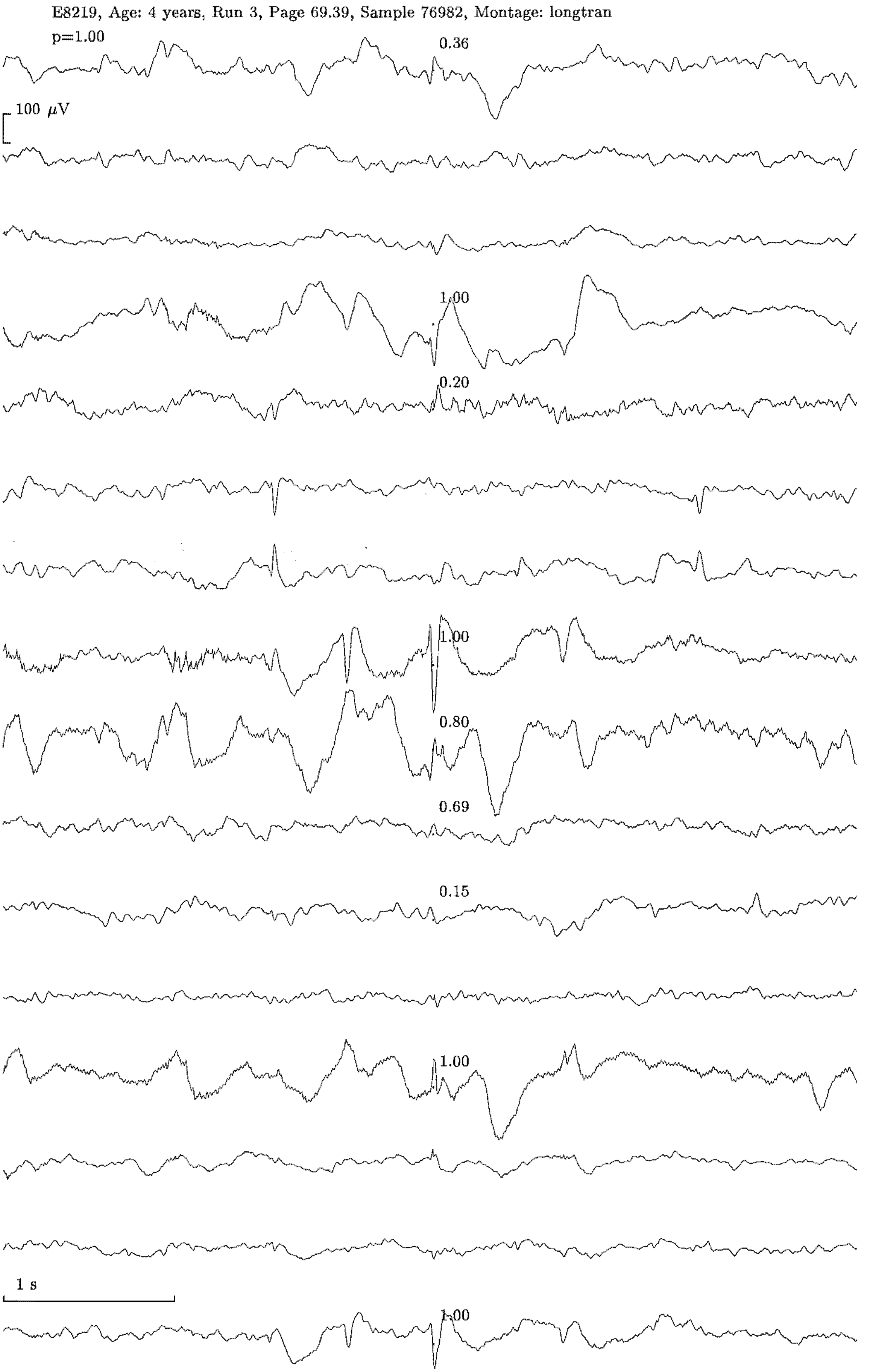

Figure A.6 Three events in EEG 12 (E8219): the first and last were labelled definite, the middle event was labelled questionable. The first event had high single-channel probabilites but just failed to qualify as a focal event. For the second event, channel eight received a definite probability but channel nine was assigned zero probability. Only the last definite ED was marked as a detection by the wavelet-based system, no definite ED was found by the Hybrid I system (p. 135).

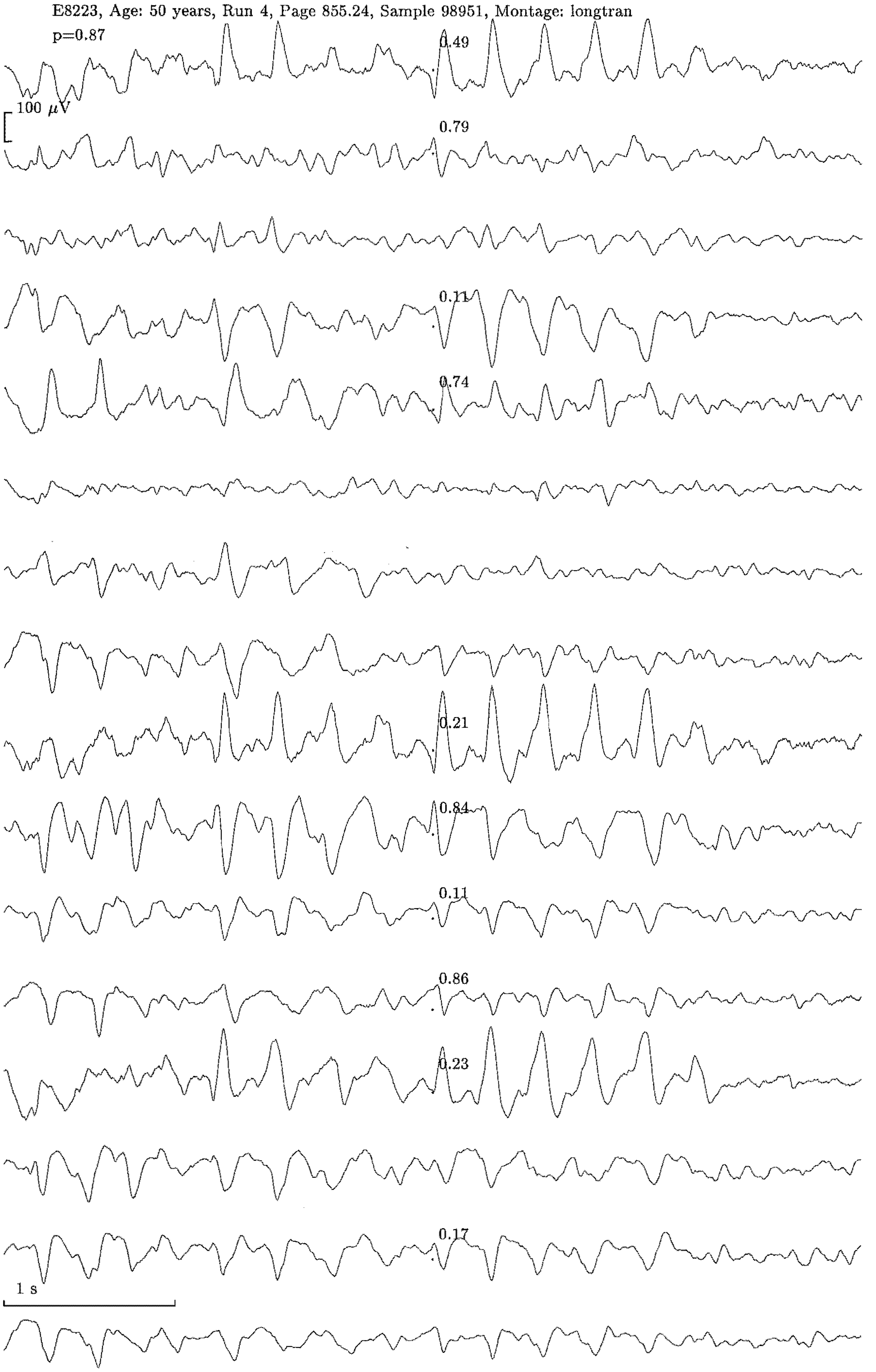

Figure A.7 This event in EEG 5 (E8223) was marked as definite, none and questionable by the three EEGers. It was detected by the Hybrid I system and the wavelet-based system (p. 138).

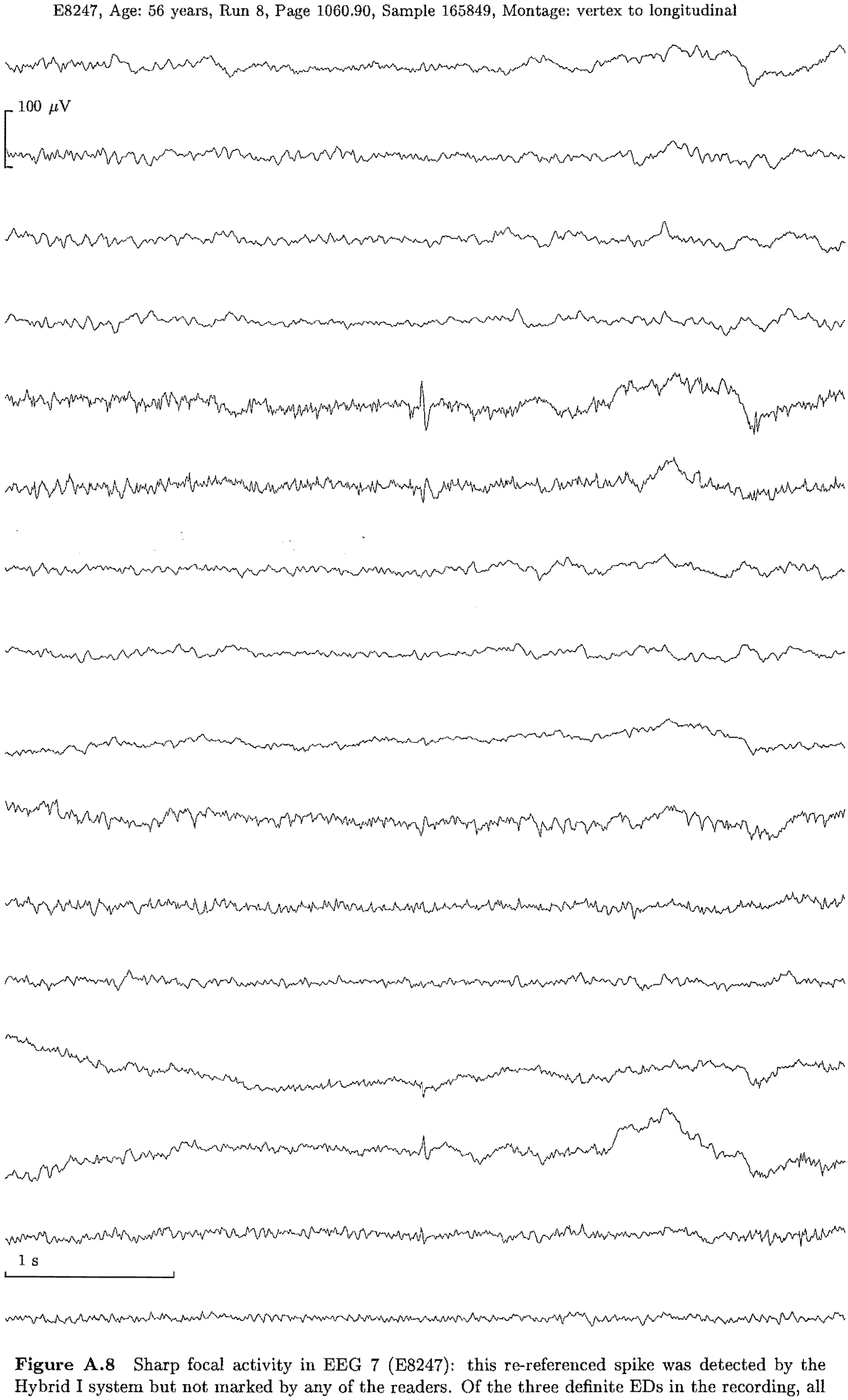

Figure A.8 Sharp focal activity in EEG 7 (E8247): this re-referenced spike was detected by the Hybrid I system but not marked by any of the readers. Of the three definite EDs in the recording, all were missed by Hybrid I and two were missed by the wavelet-based system. (p. 138).

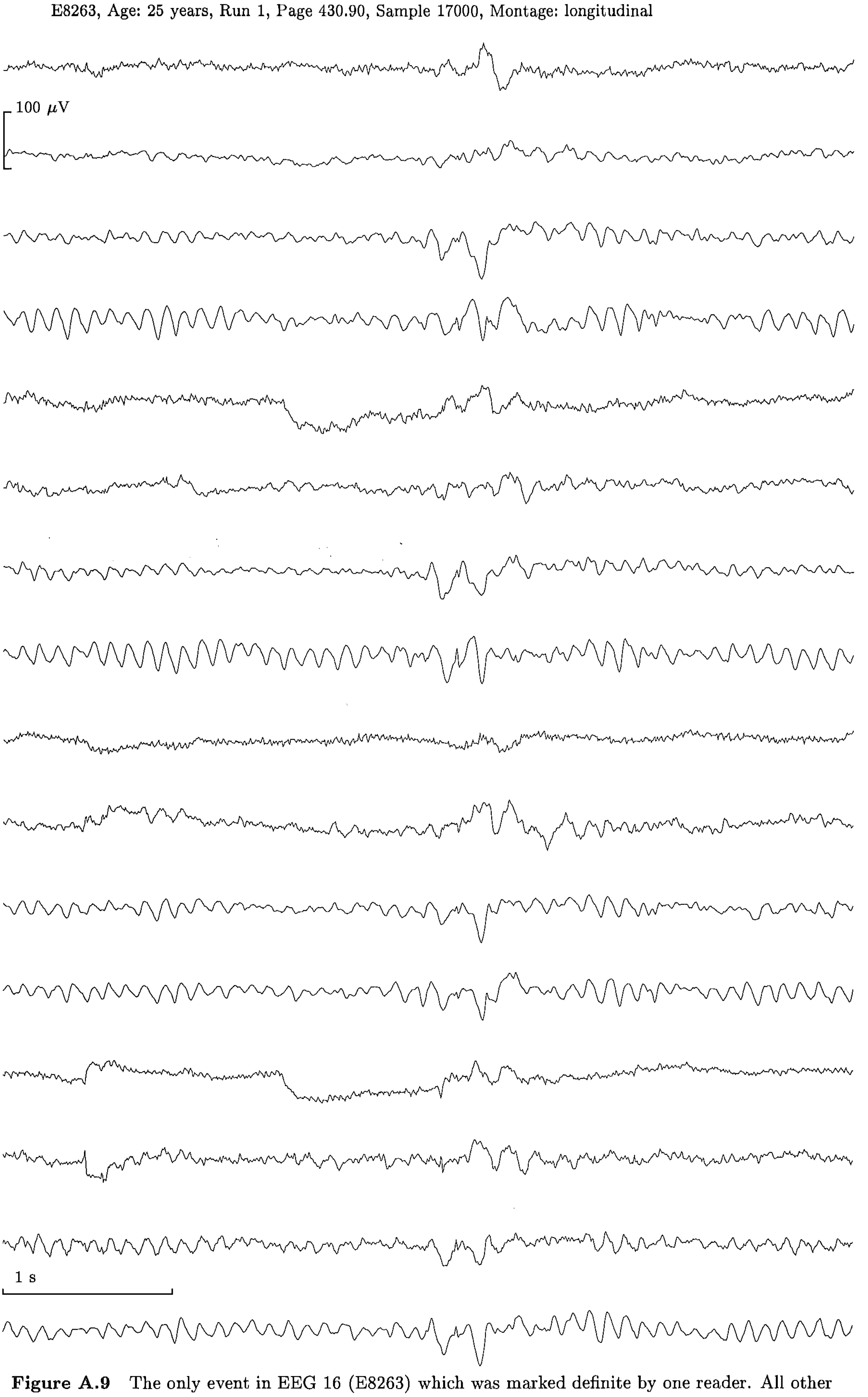

Figure A.9 The only event in EEG 16 (E8263) which was marked definite by one reader. All other events received questionable perception values. Both wavelet-based system and Hybrid I did not detect any ED in this recording (p. 145).

E8297, Age: 6 months, Run 1, Page 811.04, Sample 109792, Montage: longitudinal
p=0.55, focal upgrade

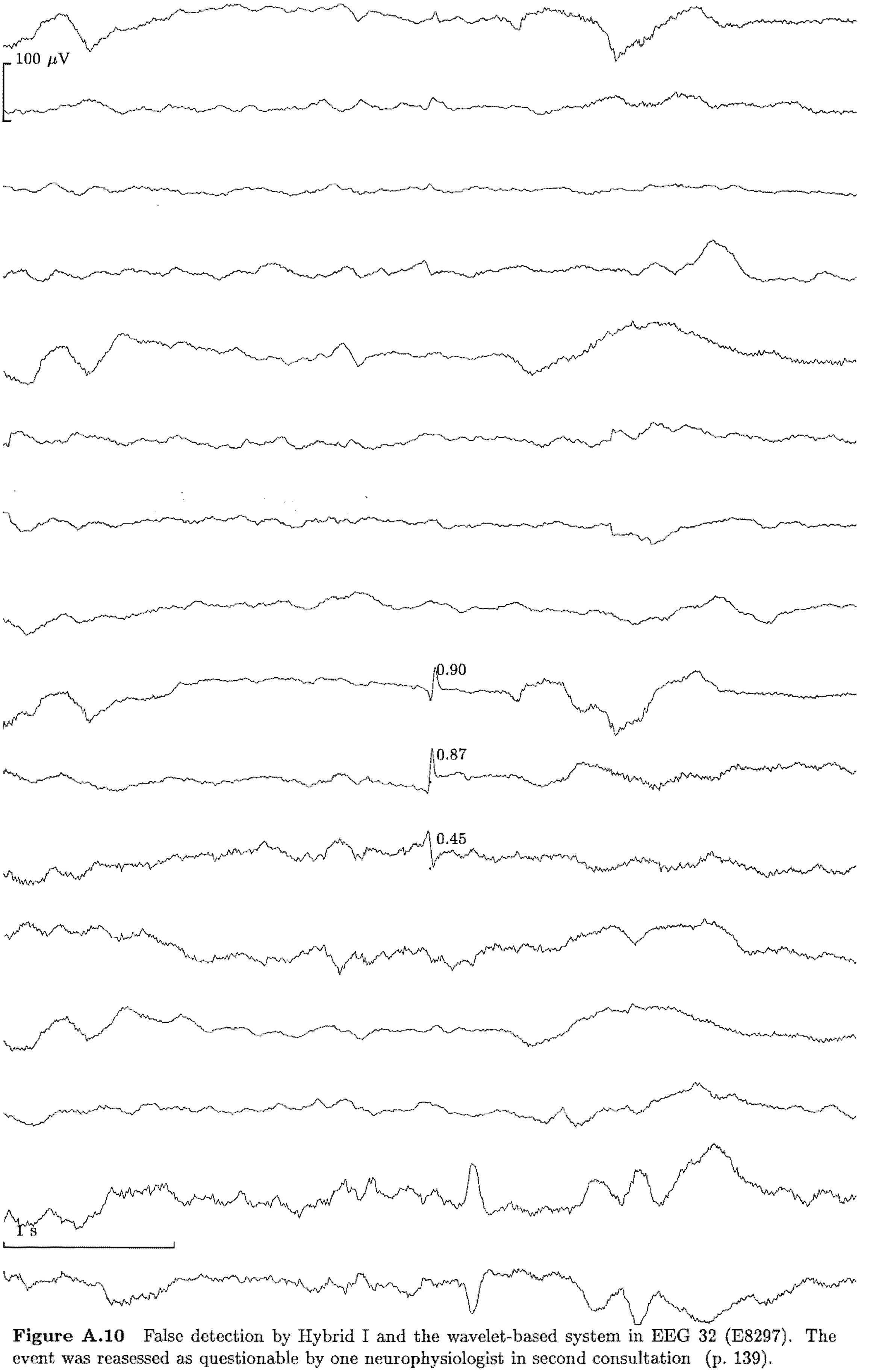

Figure A.10 False detection by Hybrid I and the wavelet-based system in EEG 32 (E8297). The event was reasessed as questionable by one neurophysiologist in second consultation (p. 139).

E8344, Age: 57 years, Run 2, Page 490.15, Sample 53500, Montage: earref to longitudinal

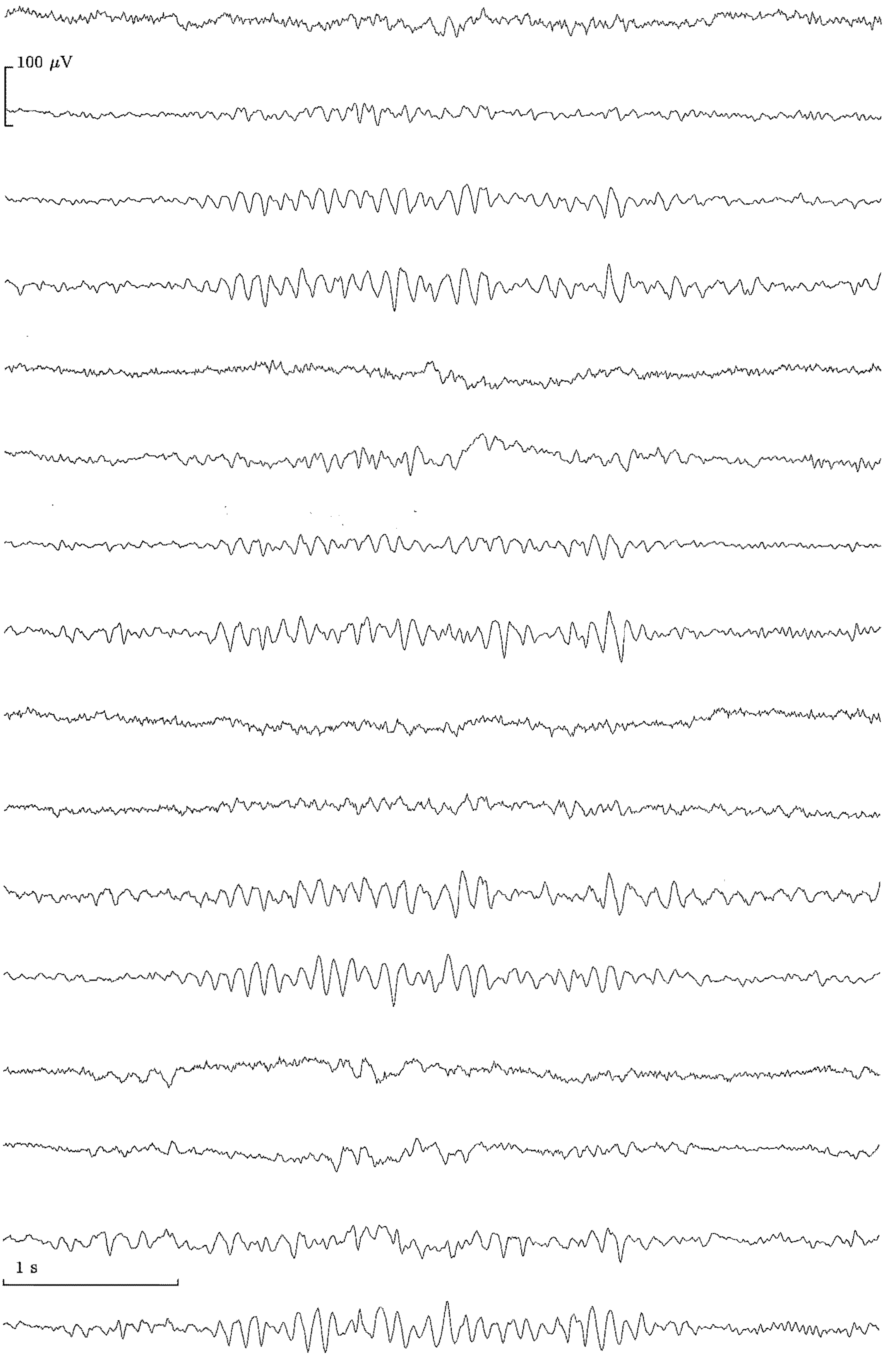

Figure A.11 Strong alpha spindles with characteristic occipital distribution (p. 7).

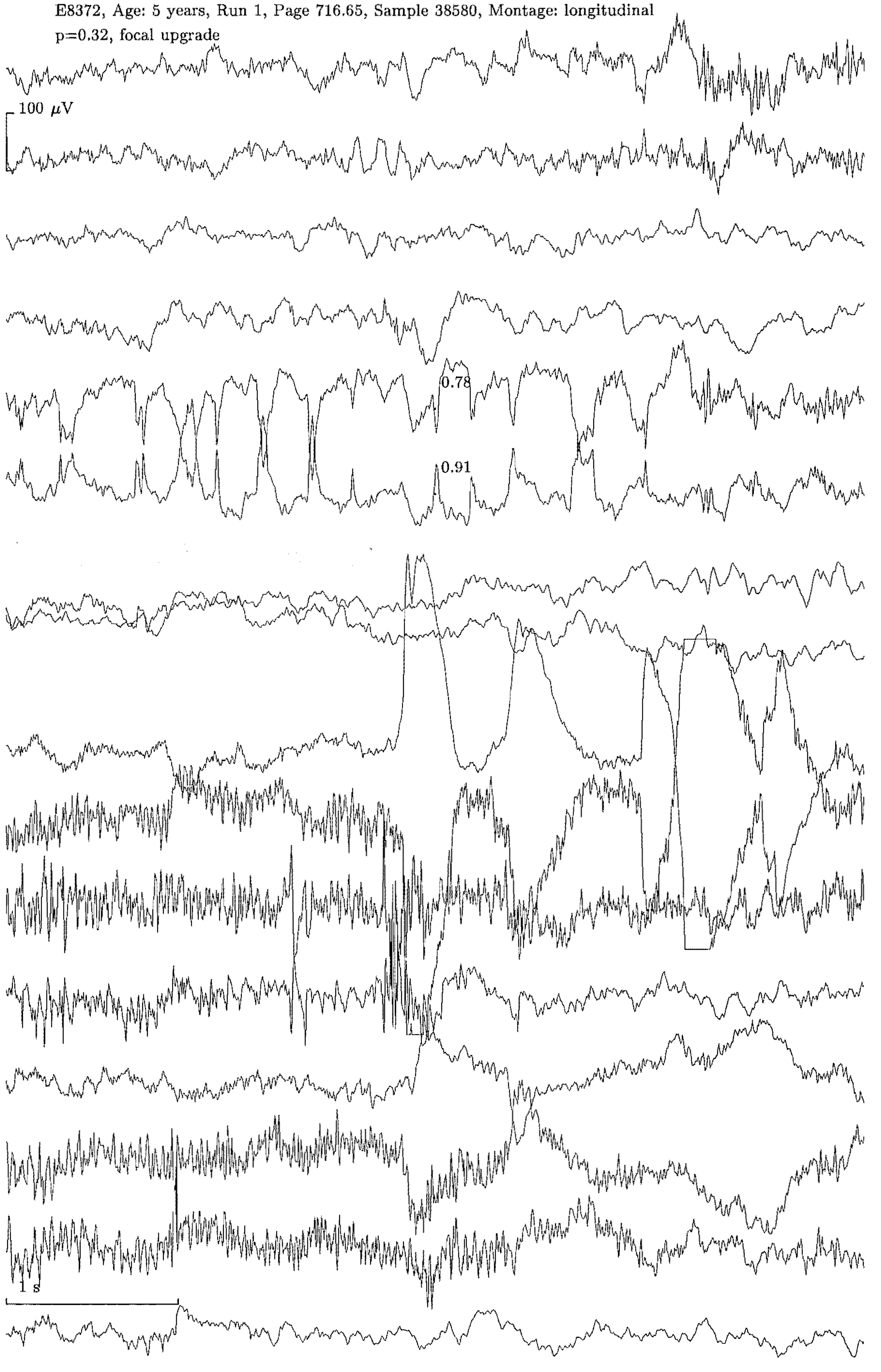

Figure A.12 Series of electrode artifacts in EEG 10 (E8372): both Hybrid I and the wavelet-based system detected one of the electrode artifacts. All five definite events elswhere in the recording were detected by the wavelet-based system and missed by Hybrid I. (p. 138).

E8380, Age: 7 months, Run 1, Page 279.66, Sample 151852, Montage: longitudinal
p=0.89

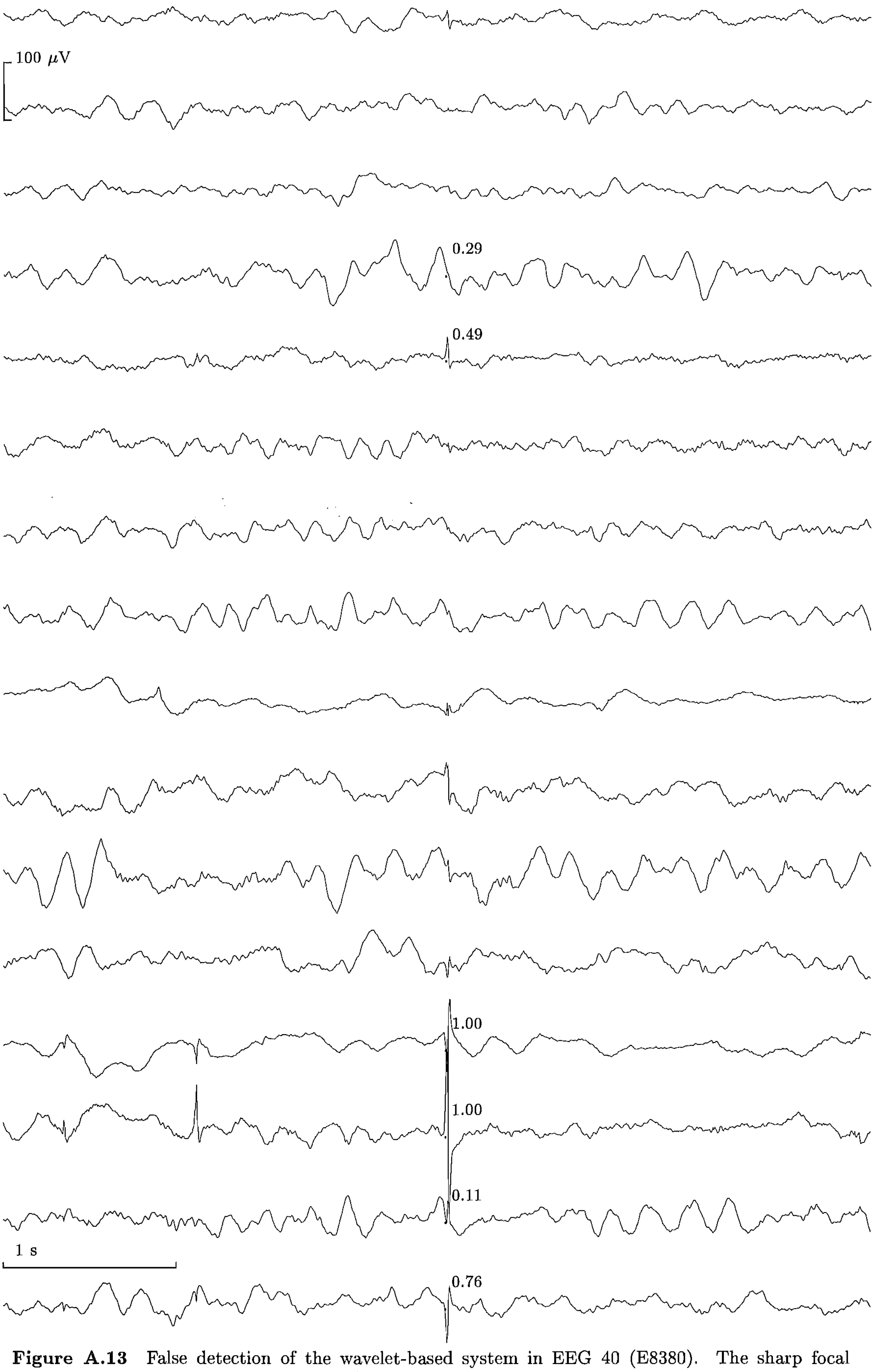

Figure A.13 False detection of the wavelet-based system in EEG 40 (E8380). The sharp focal activity was not marked as epileptiform (p. 138).

E8399, Age: 48 years, Run 10, Page 565.15, Sample 205516, Montage: longitudinal

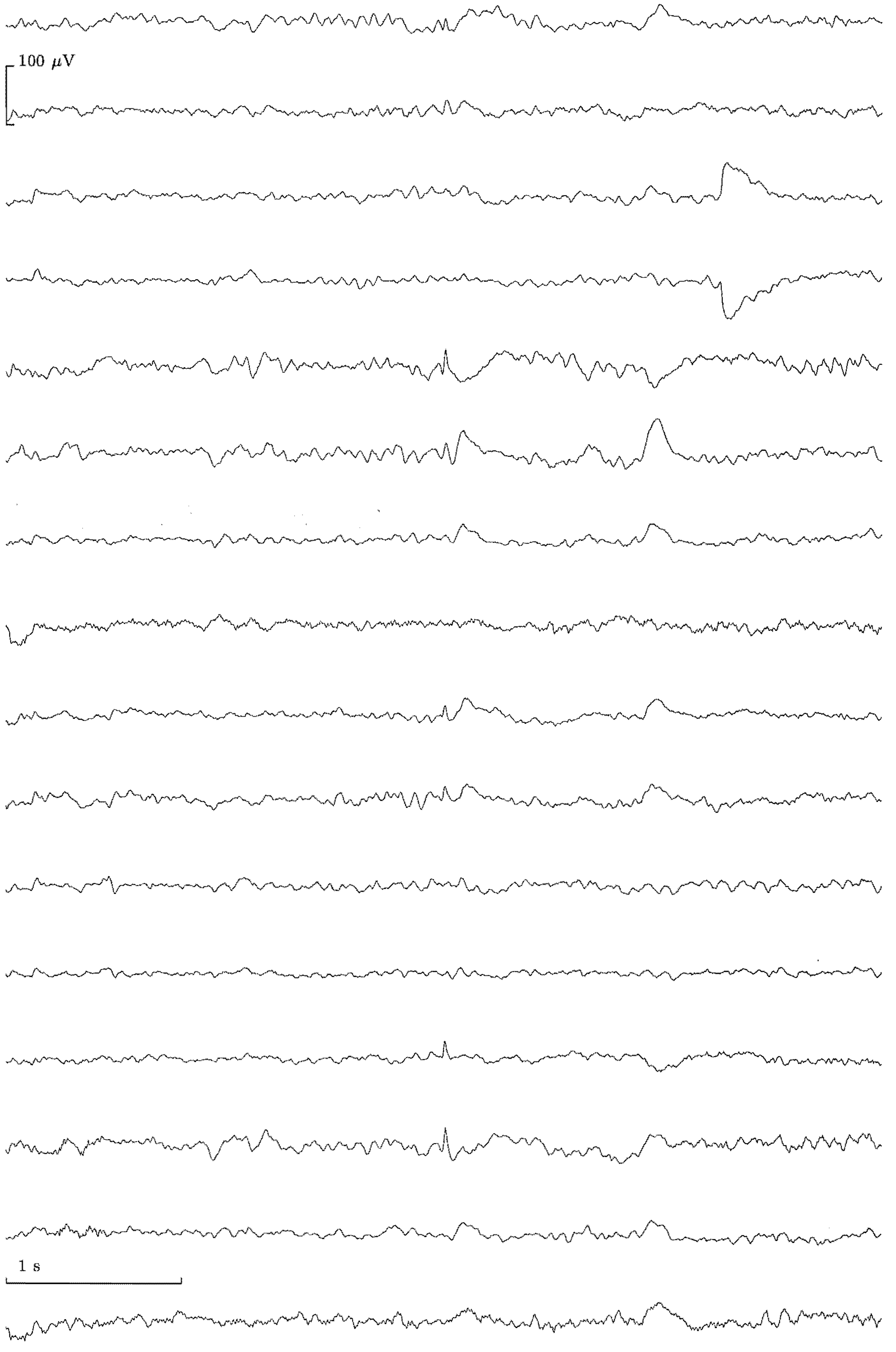

Figure A.14 False detection by Hybrid I in EEG 31 (E8399) which was labeled as questionable by one neurophysiologist in second consultation (p. 139).

E8419, Age: 43 years, Run 9, Page 187.41, Sample 244758, Montage: longitudinal

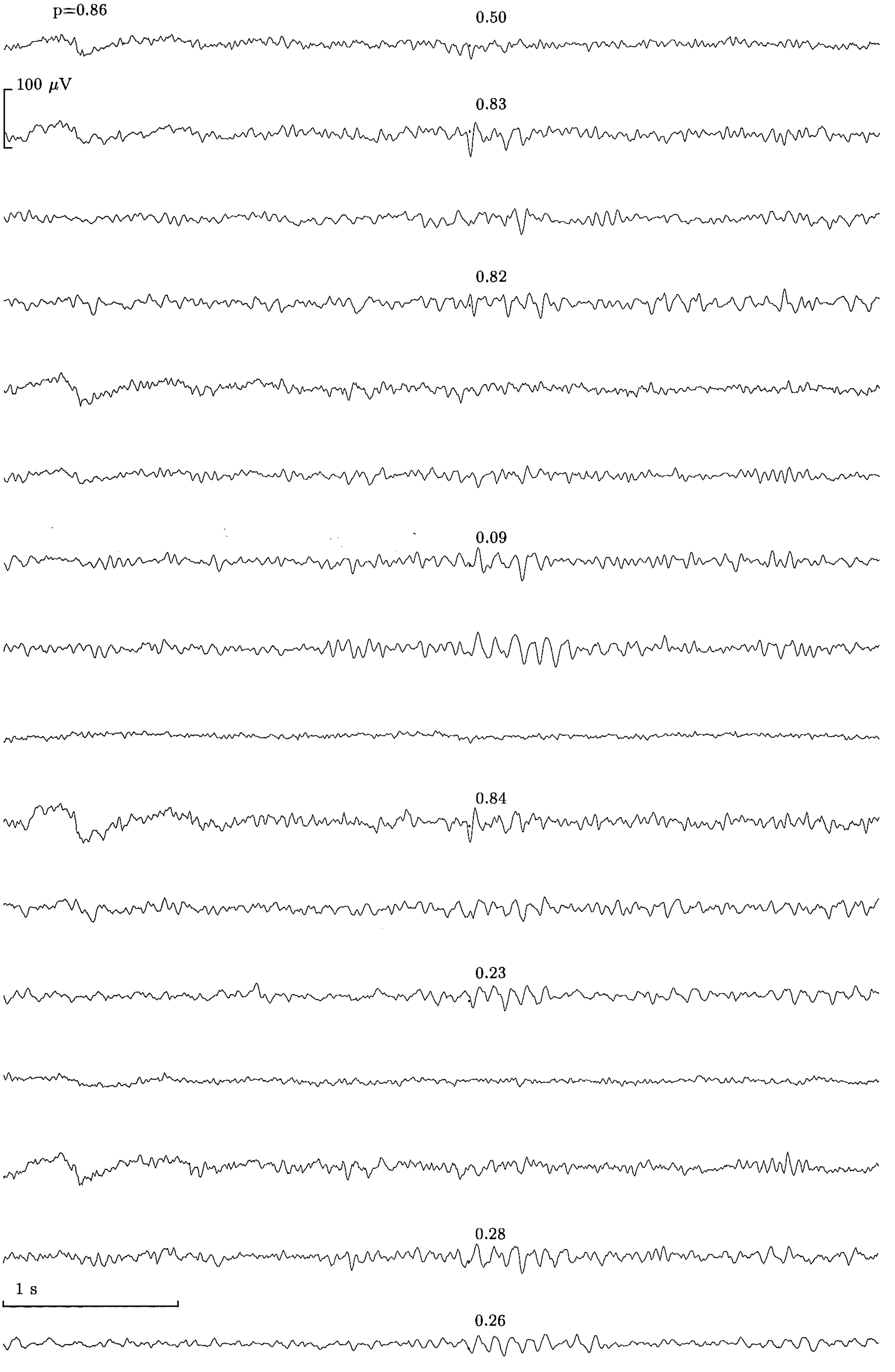

Figure A.15 False detection in EEG 42 (E8419) due to sharp alpha waves (p. 138).

E8434, Age: 8 years, Run 2, Page 358.29, Sample 53207, Montage: earref to longitudinal

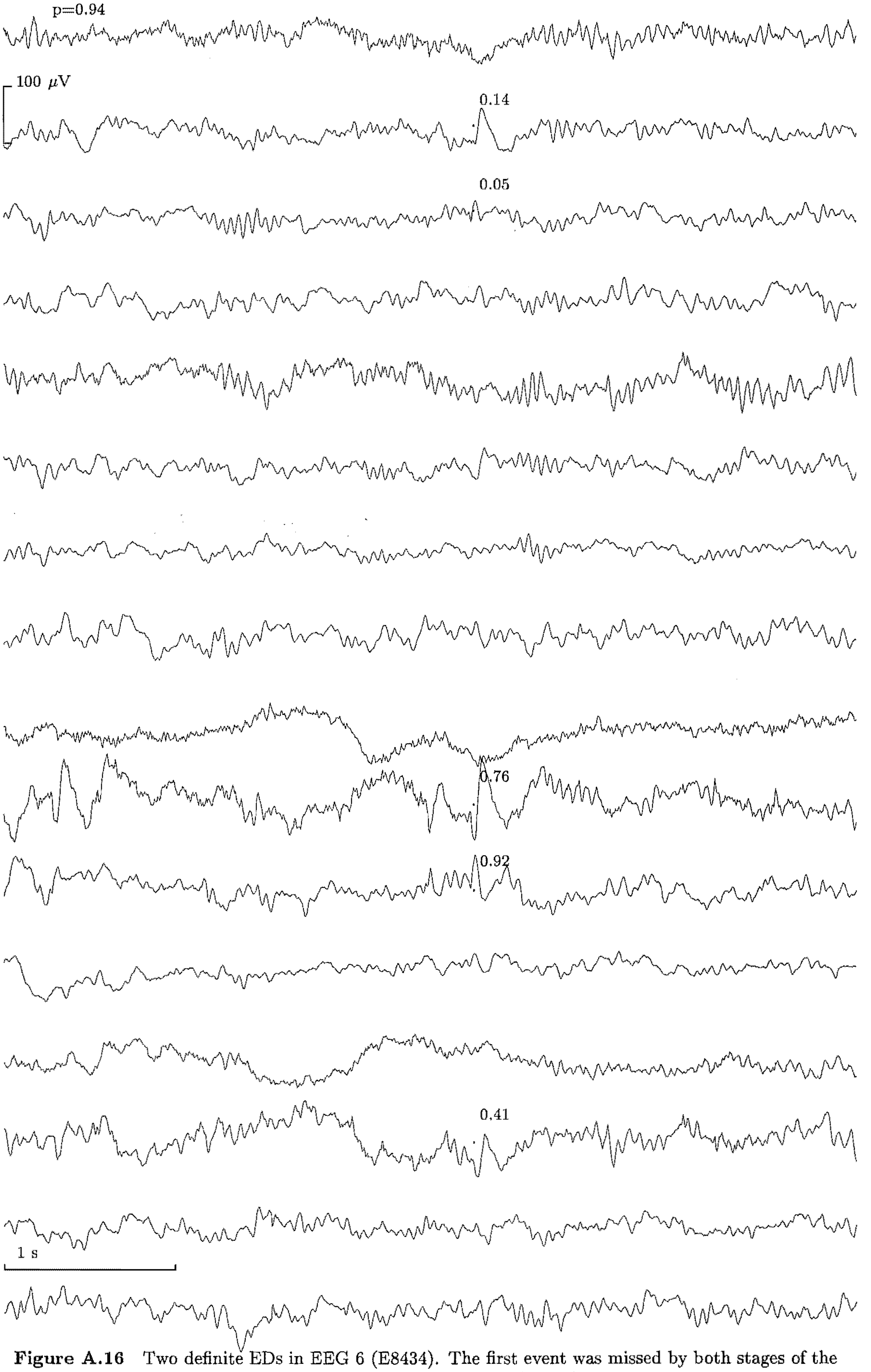

Figure A.16 Two definite EDs in EEG 6 (E8434). The first event was missed by both stages of the wavelet-based system. Both events were missed by Hybrid I. (p. 135).

E8443, Age: 25 years, Run 3, Page 251.99, Sample 67500, Montage: transver

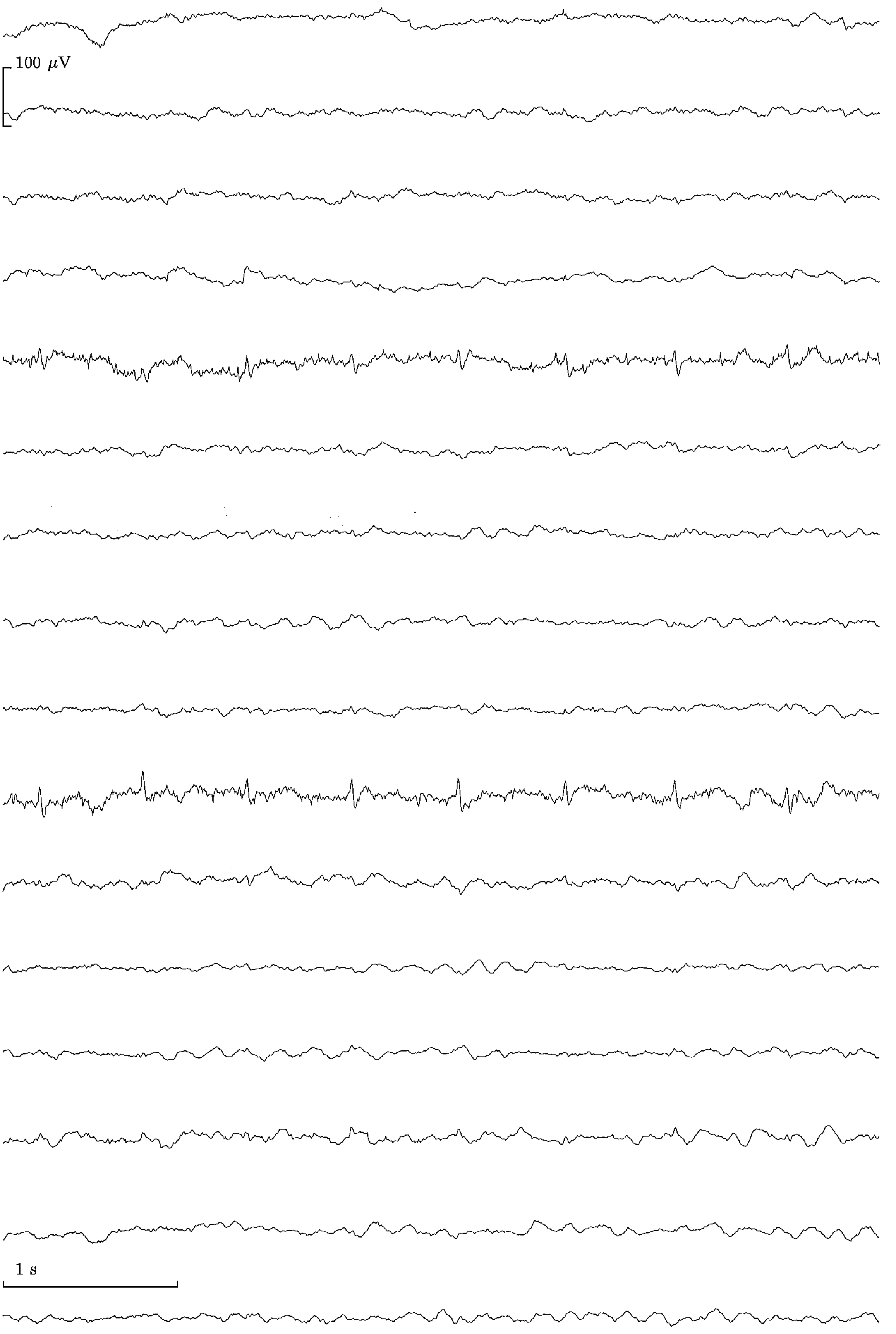

Figure A.17 ECG activity picked up by the ear reference electrodes can be seen in channels 5 and 10 (p. 10).

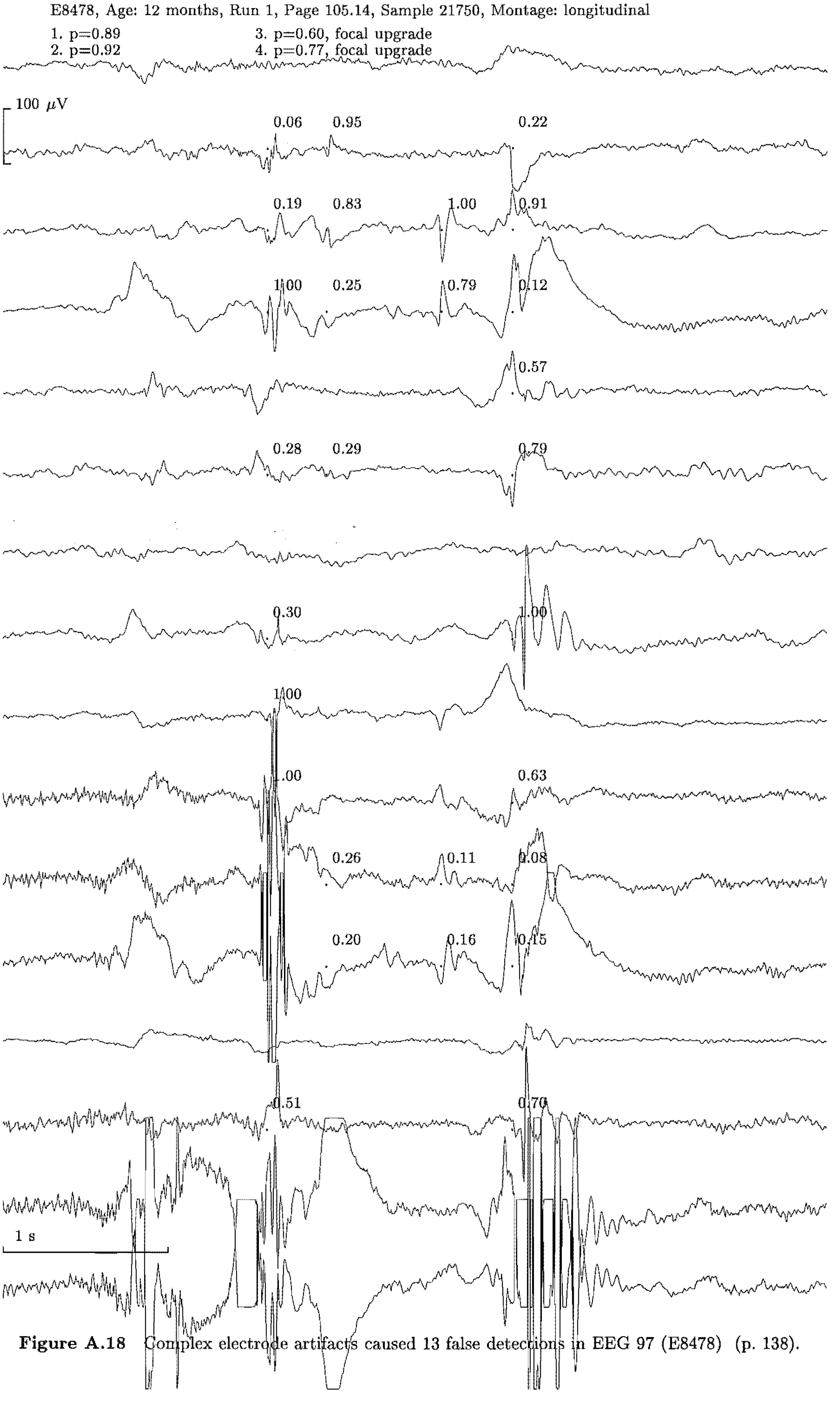

Figure A.18 Complex electrode artifacts caused 13 false detections in EEG 97 (E8478) (p. 138).

E8523, Age: 30 years, Run 10, Page 260.31, Sample 273739, Montage: longitudinal

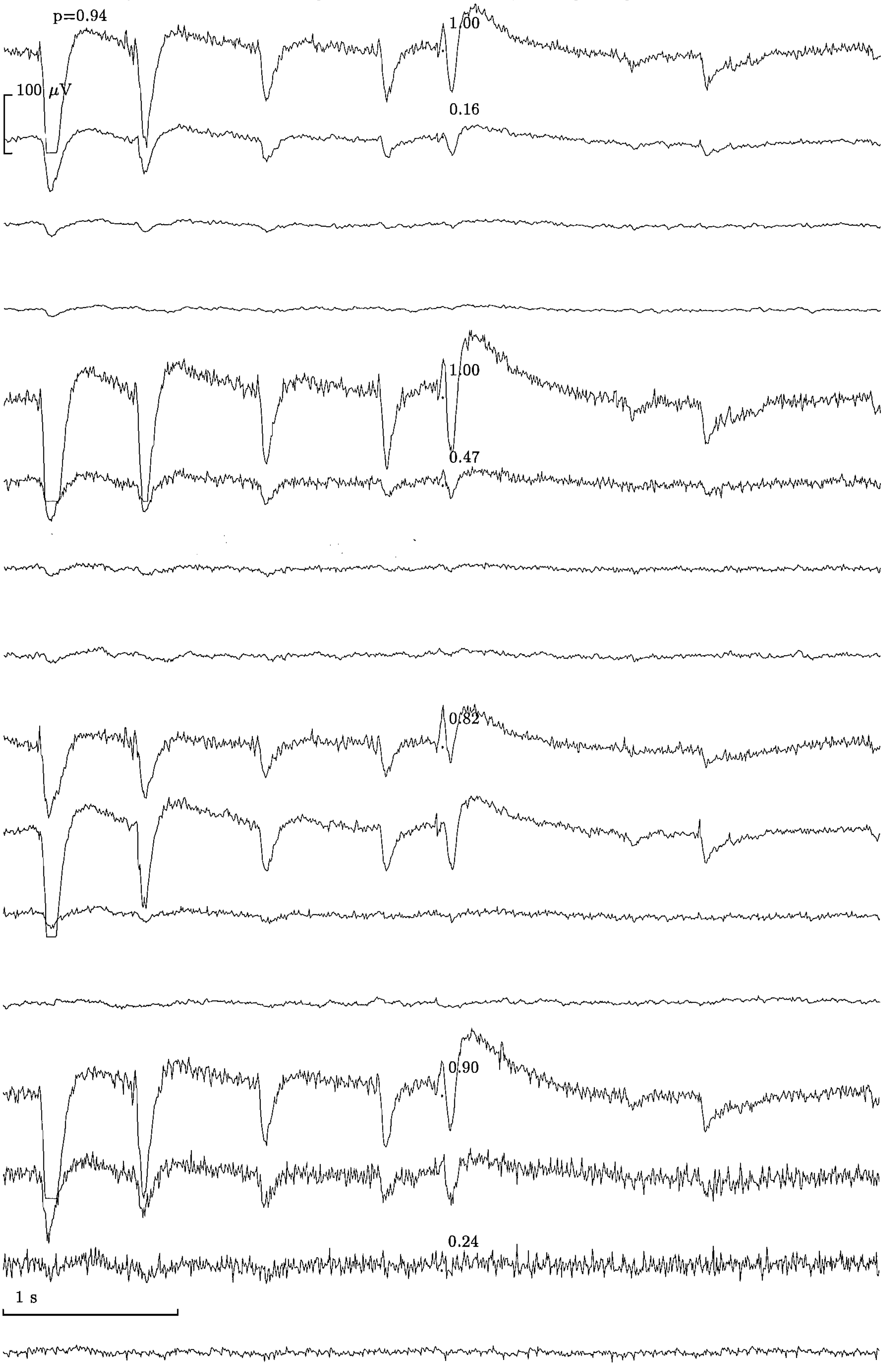

Figure A.19 An eye blink with sharp onset in EEG 29 (E8523) caused this false detection by the wavelet-based system and Hybrid I (p. 138).

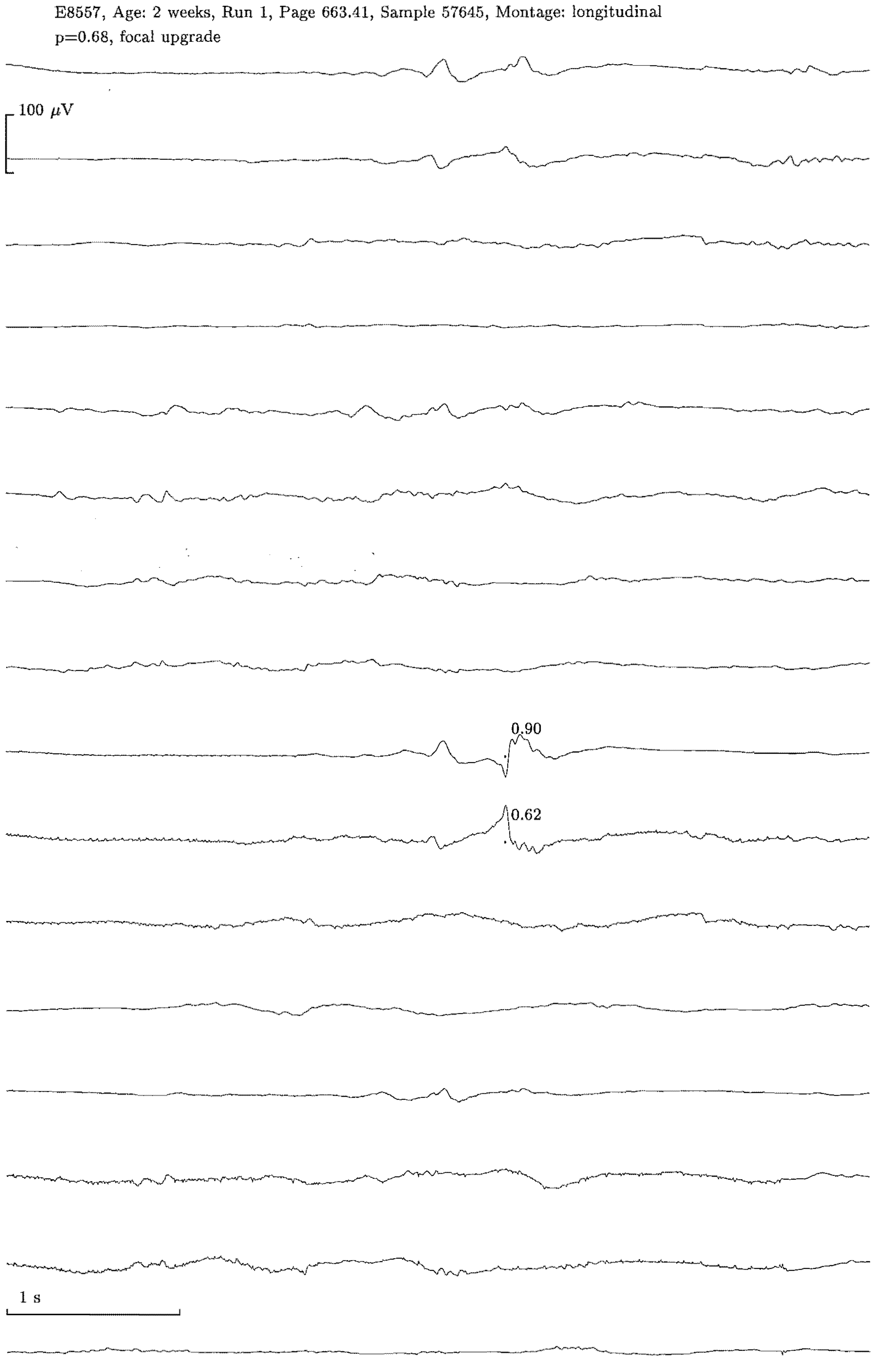

Figure A.20 Definite event in EEG 9 (E8557). This was one of five definite EDs detected by the wavelet-based system. No events were detected in this recording by Hybrid II in the most sensitive mode. (p. 135).

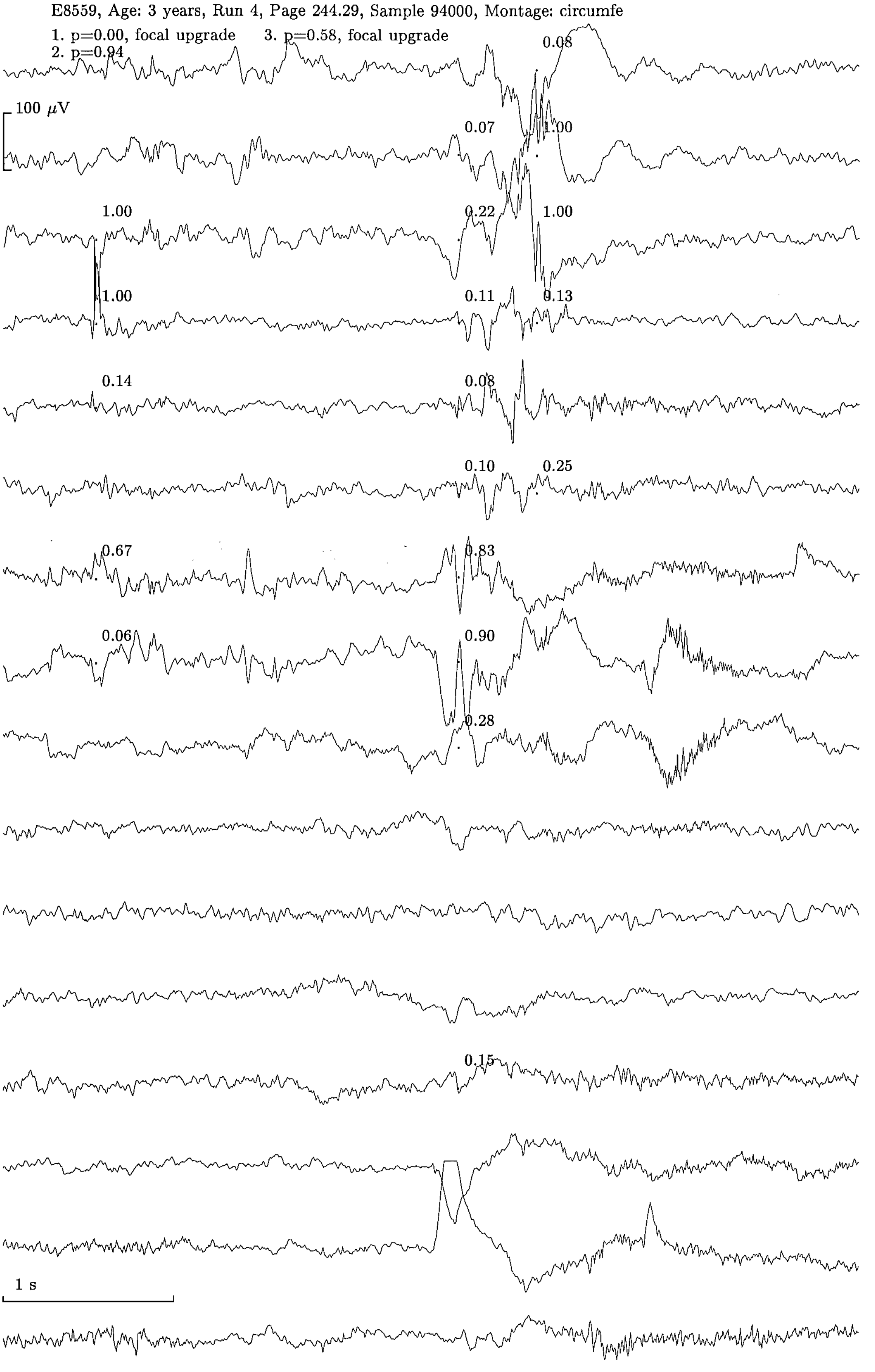

Figure A.21 Three false detections of the wavelet-based system in EEG 22 (E8559) caused by electrode and movement artifacts (p. 135).

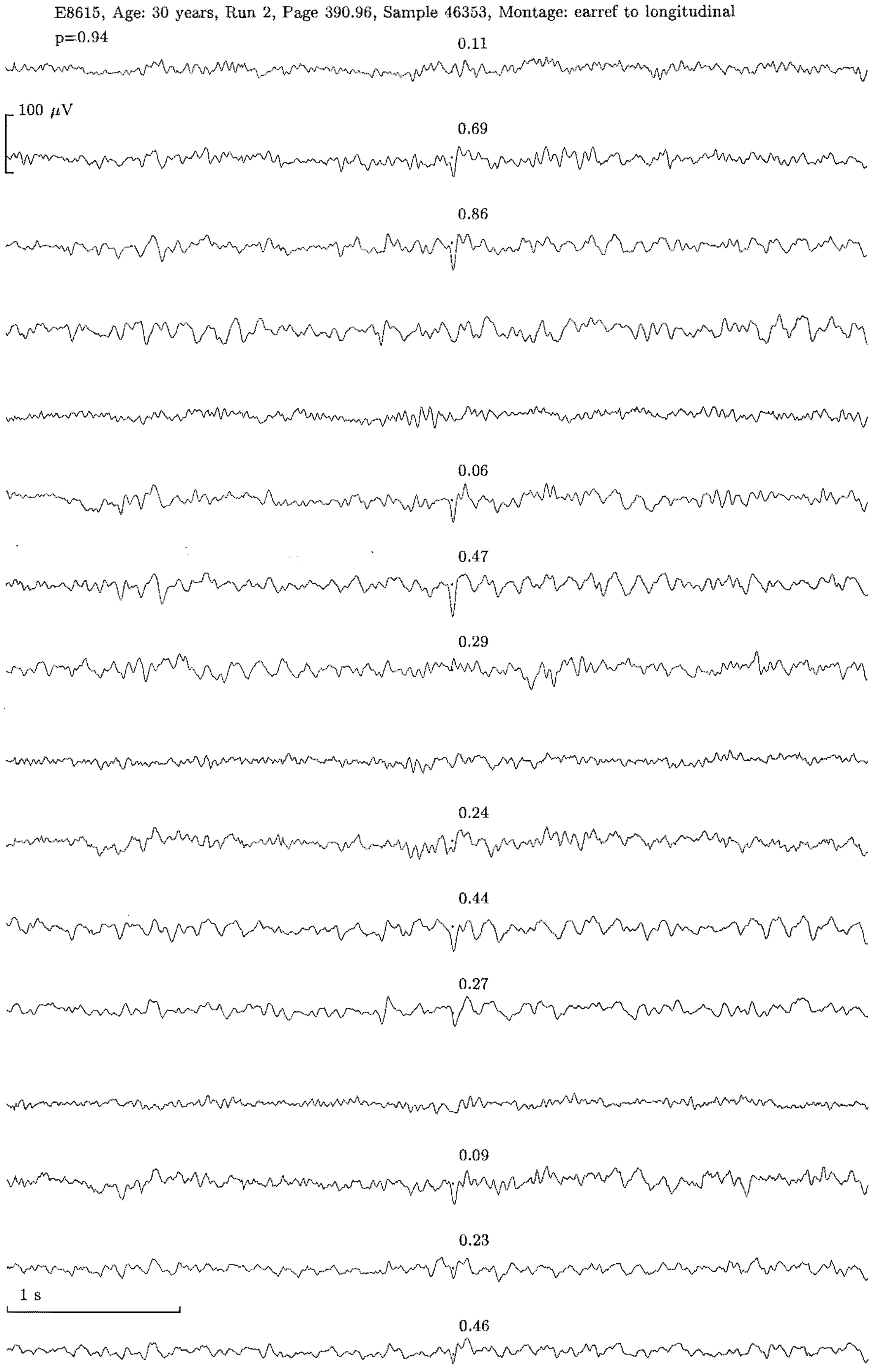

Figure A.22 False detection in EEG 70 (E8615) due to sharp alpha waves (p. 138).

REFERENCES

ANDERSEN, P. and ANDERSSON, S. (1968), *Physiological Basis of the Alpha Rhythm*, Appleton-Century-Crofts, New York.

AUGER, F., FLANDRIN, P., GONCALVES, P. and LEMOINE, O. (1997), 'Time-frequency toolbox for use with Matlab'. Reference Guide and Tutorial, Groupements de Recherche Traitement du Signal et Images, Centre National de la Recherche Scientifique, Lyon, France.

BERG, P. and SCHERG, M. (1994), 'A fast method for forward computation of multiple-shell spherical head models', *Advances in Neural Information Processing Systems*, Vol. 90, pp. 58-64.

BERGER, H. (1929), 'Über das Elektroenkephalogramm des Menschen', *Arch. Psychiat. NervKrankh.*, Vol. 87, p. 527.

BERKNER, K. and WELLS, R. (1997), 'A fast approximation to the continuous wavelet transform with applications', In *Proc. 31st Asilomar Conference on Signals, Systems and Computers. November 2-5, 1997 Pacific Grove, California*, pp. 42–45.

BERKNER, K. and WELLS, R. (1999), 'A new hierarchical scheme for approximating the continuous wavelet transform with applications to edge detection', *IEEE Signal Processing Letters*, Vol. 6, pp. 193–195.

BINNIE, C. (1993), 'Electroencephalography', In LAIDLAW, J., RICHENS, A. and CHADWICK, D. (Eds.), *A Textbook of Epilepsy*, Longman, London, pp. 277–348.

BINNIE, C., VAN EMDE BOAS, W. and TRENITE, D.K.N. (1986), 'Acute effects of lamotrigine (BW430C) in persons with epilepsy', *Epilepsia*, Vol. 27, pp. 248–254.

BIRKEMEIER, W., FONTAINE, A., CELESIA, G. and MA, K. (1978), 'Pattern recognition techniques for the detection of epileptic transients in EEG', *IEEE Transactions on Biomedical Engineering*, Vol. 25, pp. 213–217.

BLACK, M. and JONES, R. (1998), 'Sensitivity and selectivity for continuous perception values - a comment', *Electroencephalography and Clinical Neurophysiology*, Vol. 106, pp. 457–459.

BLACK, M., JONES, R., DINGLE, A. and CARROLL, G. (1997), 'An evaluation of the Christchurch automated EEG analysis system: detection of individual spikes'. Internal Report, Dept. of Medical Physics and Bioengineering, Christchurch Hospital, Christchurch, New Zealand.

BLACK, M., JONES, R., CARROLL, G., DINGLE, A., DONALDSON, I. and PARKIN, P. (2000), 'Real-time detection of epileptiform activity in the EEG: A blinded clinical trial', *Clinical Electroencephalography*, Vol. 31, pp. 122–130.

CANNY, J. (1986), 'A computational approach to edge detection', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, pp. 679–698.

CARMONA, R.A., HWANG, W.L. and TORRESANI, B. (1997), 'Characterization of signals by the ridges of their wavelet transforms', *IEEE Transactions on Signal Processing*, Vol. 45, pp. 2586–2590.

CARRIE, J. (1972a), 'A hybrid computer system for detecting and quantifying spike and wave EEG patterns', *Electroencephalography and Clinical Neurophysiology*, Vol. 41, pp. 339–341.

CARRIE, J. (1972b), 'A hybrid computer technique for detecting sharp EEG transients', *Electroencephalography and Clinical Neurophysiology*, Vol. 41, pp. 336–338.

CHADWICK, D. (1993), 'Seizures and epilepsy in adults', In LAIDLAW, J., RICHENS, A. and CHADWICK, D. (Eds.), *A Textbook of Epilepsy*, Longman, London, pp. 165–204.

CHARNIAK, E. and MCDERMOTT, D. (1985), *Introduction to Artificial Intelligence*, Addison-Wesley, Reading, Mass.

CHATRIAN, G., BERGAMINI, L., DONDEY, M., KLAAS, D., LENNOX-BUCHTHAL, M. and PETERSEN, I. (1977), 'A glossary of terms most commonly used by clinical electroencephalographers', *Electroencephalography and Clinical Neurophysiology*, Vol. 37, pp. 538–548.

CHOI, H. and WILLIAMS, W. (1989), 'Improved time-frequency representation of multicomponent signals using exponential kernels', *IEEE Transactions on Acoustics, Speech, and Signal Processing*, pp. 862–871.

CLARENCON, D., RENAUDIN, M., GOURMELON, P., KERCKHOEVE, A., CATERINI, R., BOIVIN, E., ELLIS, P., HILLE, B. and FATOME, M. (1996), 'Real-time spike detection in EEG signals using the wavelet transform and a dedicated digital signal processor card', *Electroencephalography and Clinical Neurophysiology*, Vol. 70, pp. 5–14.

COHEN, A., DAUBECHIES, I. and FEAUVEAU, J. (1992), 'Biorthogonal basis of compactly supported wavelets', *Communications on Pure and Applied Mathematics*, Vol. 45, pp. 485–560.

Commission on Classification and Terminology of the International League Against Epilepsy (1981), 'Proposal for the revised clinical and electroencephalographic classification of epileptic seizures', *Epilepsia*, Vol. 22, pp. 489–501.

Commission on Classification and Terminology of the International League Against Epilepsy (1989), 'Proposal for the revised classification of epilepsies and epileptic syndromes', *Epilepsia*, Vol. 30, pp. 389–399.

CUFFIN, B. and COHEN, D. (1979), 'Comparison of the magnetoencephalogram and the electroencephalogram', *Electroencephalography and Clinical Neurophysiology*, pp. 132–146.

DA SILVA, F.H.L., DIJK, A. and SMITS, H. (1975), 'Detection of nonstationarities in EEGs using the autoregressive model — an application to EEGs of epileptics', In (Ed.), *CEAN : Computerized EEG Analysis*, Gustav Fischer Verlag, Stuttgart, pp. 180–199.

DAMON, R. and HARVEY, W. (1987), *Experimental Design, ANOVA, and Regression*, Harper & Row, New York.

D'ATTELIS, C., ISAACSON, S. and SIRNE, R. (1997), 'Detection of epileptic events in electroencephalograms using wavelet analysis', *Annals of Biomedical Engineering*, Vol. 25, pp. 286–293.

DAUBECHIES, I. (1988), 'Orthonormal bases of compactly supported wavelets', *Communications on Pure and Applied Mathematics*, Vol. 41, pp. 909–996.

DAUBECHIES, I. (1990), 'The wavelet transform, time-frequence localization and signal analysis', *IEEE Transactions on Information Theory*, Vol. 36, pp. 961–1005.

DAUBECHIES, I. (1992), *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

DAVEY, B.L.K., FRIGHT, W.R., CARROLL, G.J. and JONES, R.D. (1989), 'Expert system approach to detection of epileptiform activity in the EEG', *Medical and Biological Engineering and Computing*, Vol. 27, pp. 365–370.

DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (1996), *A Probabilistic Theory of Pattern Recognition*, Springer-Verlag, New York.

DINGLE, A. (1992), *Engineering in Brain Research: Processing Electroencephalograms and Chaos in Neural Networks*, PhD thesis, University of Canterbury, Christchurch, New Zealand.

DINGLE, A., JONES, R., CARROLL, G. and FRIGHT, W. (1993), 'A multistage system to detect epileptiform activity in the EEG', *IEEE Transactions on Biomedical Engineering*, Vol. BME-40, pp. 1260–1268.

DUFFY, F.H., IYER, V.G. and SURWILLO, W.W. (1989), *Clinical Electroencephalography and Topographic Brain Mapping*, Springer Verlag.

DÜMPELMANN, M. and ELGER, C.E. (1999), 'Visual and automatic investigation of epileptiform spikes in intracranial EEG recordings', *Epilepsia*, Vol. 40, pp. 275–285.

DURKA, P.J. and BLINOWSKA, K.J. (1996), 'Matching pursuit parameterization of sleep spindels', In *Proc. 18th Ann. Int. Conf. of IEEE Engineering in Medicine and Biology Society, 31 Oct–3 Nov, Amsterdam.*

EBERHART, R., DOBBINS, R. and WEBBER, W. (1989), 'Neural network design for EEG spike detection', In *Proc. 15th Notheast Bioengineering Conference*, pp. 97–98.

ECKHORN, R., BAUER, R., JORDAN, W., BROSCH, M., KRUSE, W., MUNK, M. and REITBOECK, H. (1988), 'Coherent oscillations: A mechanism of feature linking in the visual cortex?', *Biological Cybernetics*, Vol. 60, pp. 121–130.

EDLINGER, G., WACH, P. and PFURTSCHELLER, G. (1998), 'On the realization of an analytic high-resolution EEG', *IEEE Transactions on Biomedical Engineering*, Vol. 45, pp. 736–745.

EGAN, J. (1975), *Signal Detection Theory and ROC Analysis*, Academic Press, New York.

EHRENBERG, B. and PENRY, K. (1976), 'Computer recignition of generalized spike-wave discharges', *Electroencephalography and Clinical Neurophysiology*, Vol. 41, pp. 25–36.

EVERTSZ, C., BERKNER, K. and BERGHORN, W. (1995), 'A local multiscale characterization of edges applying the wavelet transform', In *Proc. Nato A.S.I., Fractal Image Encoding and Analysis, Trondheim*, July, pp. 42–54.

FERGUSON, A. and STROINK, G. (1997), 'Factors affecting the accuracy of the boundary element method in the forward problem - i. calculating surface potentials', *IEEE Transactions on Biomedical Engineering*, Vol. 44, pp. 1139–1155.

FISCH, B. (1991), *Spehlmann's EEG Primer*, Elsevier, Amsterdam.

FISHER, R. (1936), 'The use of multiple measurements in taxonomic problems', *Annals of Eugenics*, Vol. 7, pp. 179–188.

FISHER, R. (1938), 'The statistical utilization of multiple measurements', *Annals of Eugenics*, Vol. 8, pp. 376–386.

FRANASZCZUK, P., BERGEY, G., DURKA, P. and EISENBERG, H. (1998), 'Time-frequency analysis using the matching pursuit algorithm applied to seizures originating from the mesial temporal lobe', *Electroencephalography and Clinical Neurophysiology*, Vol. 106, pp. 513–521.

FRISCH, M. and MESSER, H. (1992), 'The use of the wavelet transform in the detection of an unknown transient signal', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, pp. 892–897.

FU, L. (1994), 'Rule generation from neural networks', *IEEE Transactions on Systems Man and Cybernetics*, Vol. 24, pp. 1114–1124.

FUCHS, M., DRENCKHAHN, R., WISCHMANN, H.A. and WAGNER, M. (1998), 'An improved boundary element method for realistic volume-conductor modeling', *IEEE Transactions on Biomedical Engineering*, Vol. 45, pp. 980–997.

GABOR, A.J. and SEYAL, M. (1992), 'Automated interictal EEG spike detection using artificial neural networks', *Electroencephalography and Clinical Neurophysiology*, Vol. 83, pp. 271–280.

GABOR, A.J., LEACH, R.R. and DOWLA, F.U. (1996), 'Automated seizure detection using a self-organized neural networks', *Electroencephalography and Clinical Neurophysiology*, Vol. 99, pp. 257–266.

GEVA, A. and KEREM, D. (1998), 'Forecasting generalized epileptic seizures from the EEG signal by wavelet analysis and dynamic unsupervised fuzzy clustering', *IEEE Transactions on Biomedical Engineering*, Vol. 45, pp. 1205–1211.

GHAHREMANI, D., MAKEIG, S., JUNG, T.P., BELL, A. and SEJNOWSKI, T. (1996), 'Independent component analysis of simulated EEG using three-shell spherical head model', *Institute for Neural Computation Technical Report*, Vol. 1, pp. 1–22.

GLOVER, J., KTONAS, P., RAGHAVAN, N., URUNELA, J., VELAMURI, S. and REILLY, E. (1986), 'A multichannel signal processor for the detection of epileptogenic sharp transients in the EEG', *IEEE Transactions on Biomedical Engineering*, Vol. 33, pp. 1121–1128.

GLOVER, J.R., RAGHAVAN, N., KTONAS, P.Y. and FROST, J. (1989), 'Context-based automated detection of epileptogenic sharp transients in the EEG: Elimination of false positives', *IEEE Transactions on Biomedical Engineering*, Vol. 36, pp. 519–527.

GOELZ, H., JONES, R. and BONES, P. (2000a), 'A new family of fast wavelet filters for transient detection'. submitted to Applied and Computational Harmonic Analysis.

GOELZ, H., JONES, R. and BONES, P. (2000b), 'Wavelet analysis of transient biomedical signals and its application to detection of epileptiform activity in the EEG', *Clinical Electroencephalography*, Vol. 41, pp. 181–191.

GOLDMAN, M. (1973), *Principles of Clinical Electrocardiography*, Lange Medical Publications, Los Altos, California.

GÖLZ, H. (1995), *Transiente raum-zeitliche Korrelationen elektrischer Hirnsignale (Transient Spatio-Temporal Correlations of Electrical Brain Signals)*, Master's thesis, Philipps University, Marburg, Germany. [German Diplomarbeit].

GÖLZ, H., RÖSLER, F., RÖDER, B. and HENNIGHAUSEN, E. (1997), 'Segregation of spatiotemporal components of auditory Event Related Potentials', In WITTE, H., ZWIENER, U., SCHACK, B. and DOERING, A. (Eds.), *Proc. Third International Hans Berger Congress: Quantitative and Topological EEG and MEG analysis*, Friedrich-Schiller-Universität Jena, Germany, pp. 238–241.

GÖLZ, H., JONES, R. and BONES, P. (1999), 'Continuous wavelet transform for the detection and classification of epileptiform activity in the EEG', In *Proc. 21th Ann. Int. Conf. of IEEE Engineering in Medicine and Biology Society, 13–16 Oct, Atlanta*, p. 941.

GOTMAN (1982), 'Automatic recognition of epileptic seizure in the EEG', *Electroencephalography and Clinical Neurophysiology*, Vol. 54, pp. 530–540.

GOTMAN, J. (1985), 'Automatic recognition of interictal spikes', *Electroencephalography and Clinical Neurophysiology*, Vol. 37, pp. 93–114.

GOTMAN, J. and GLOOR, P. (1976), 'Automatic recognition and quantification of interictal epileptiform activity in the human scalp EEG', *Electroencephalography and Clinical Neurophysiology*, Vol. 41, pp. 513–529.

GOTMAN, J. and WANG, L. (1991), 'State dependent spike detection: concepts and preliminary results', *Electroencephalography and Clinical Neurophysiology*, Vol. 79, pp. 11–19.

GOTMAN, J. and WANG, L. (1992), 'State-dependent spike detection: validation', *Electroencephalography and Clinical Neurophysiology*, Vol. 83, pp. 12–18.

GREEN, R. (1995), *Detection of Epileptiform Transients in the Electroencephalogram: Real-Time Processing and Eigenspikes in Neural Networks*, Master's thesis, University of Canterbury, Christchurch, New Zealand.

GROSSMANN, A., KRONLAND-MARTINET, R. and MORLET, J. (1989), 'Reading and understanding continuous wavelet transforms', In COMBES, J.M., GROSSMANN, A. and TCHAMITCHIAN, P. (Eds.), *Wavelets: Time-Frequency Methods and Phase Space*, Springer Verlag, New York, pp. 2–20.

HAAR, A. (1910), 'Zur Theorie der orthogonalen Funktionen-Systeme', *Math. Ann.*, Vol. 69, pp. 331–371.

HARA, J., MUSHA, T. and SHANKLE, W. (1999), 'Approximating dipoles from human EEG activity: The effect of dipole source configuration on dipolarity using single dipole models', *IEEE Transactions on Biomedical Engineering*, Vol. 46, pp. 125–129.

HENNIGHAUSEN, E., HEIL, M. and RÖSLER, F. (1993), 'A correction method for DC drift artifacts', *Electroencephalography and Clinical Neurophysiology*, Vol. 86, pp. 199–204.

HERCULANO-HOUZEL, S., MUNK, M., NEUENSCHWANDER, S. and SINGER, W. (1999), 'Precisely synchronized oscillatory firing patterns require electroencephalographic activation', *Journal of Neuroscience*, Vol. 19, pp. 3992–4010.

HO, K. and CHAN, Y. (1999), 'Filter design and comparison for two fast CWT algorithms', *IEEE Transactions on Signal Processing*, Vol. 47, pp. 3013–3026.

HUBEL, D. and WIESEL, T. (1977), 'Functional architecture of macaque monkey visual cortex', *Proceedings of the Royal Society, London*, Vol. 198, pp. 1–59.

IASEMIDIS, L., PAPPAS, K., PRINCIPE, J. and SACKELLARES, J. (1996), 'Spatiotemporal dynamics of human epileptic seizures', In *Proceedings of the 3rd Experimental Chaos Conference*, pp. 26–30.

JAMES, C. (1997), *Detection of Epileptiform Activity in the Electroencephalogram using Artificial Neural Networks*, PhD thesis, University of Canterbury, Christchurch, New Zealand.

JAMES, C.J., JONES, R.D., BONES, P.J. and CARROLL, G.J. (1996a), 'The self-organizing feature map in the detection of epileptiform transients in the EEG', In *Proc. 18th Ann. Int. Conf. of IEEE Engineering in Medicine and Biology Society, 31 Oct–3 Nov, Amsterdam.*

JAMES, C.J., HAGAN, M.T., JONES, R.D., BONES, P.J. and CARROLL, G.J. (1996b), 'Spatio-temporal filtering of the EEG via neural network based multireference adaptive noise cancelling', In *Proc. 3rd New Zealand Conf. of Postgrad. Students in Engineering and Technology, 1–2 Jul 1996, Christchurch.*

JAMES, C.J., HAGAN, M.T., JONES, R.D., BONES, P.J. and CARROLL, G.J. (1997), 'Multireference adaptive noise cancelling applied to the EEG', *IEEE Transactions on Biomedical Engineering*, Vol. 44, pp. 775–779.

JAMES, C., JONES, R., BONES, P. and CARROLL, G. (1999), 'Detection of epileptiform discharges in the EEG by a hybrid system comprising mimetic, self-organized artificial neural network, and fuzzy logic stages', *Clinical Neurophysiology*, Vol. 110, pp. 2049–2063.

JASPER, H.H. (1958), 'Report of the committee of methods of clinical examination in electroencephalography', *EEG Journal*, Vol. 10, p. 370.

JONES, R., DINGLE, A., CARROLL, G., SATHERLEY, B., MUIR, S., DONALDSON, I., PARKIN, P. and BONES, P. (1994), 'A PC-based system for automated analysis of the EEG', In *Proc. 1st Medical Engineering Week of the World Conference, Taipei, Taiwan*, pp. 153–157.

JONES, R.D., DINGLE, A.A., CARROLL, G.J., GREEN, R.D., BLACK, M., DONALDSON, I.M., PARKIN, P.J., BONES, P.J. and BURGESS, K.L. (1996), 'A system for detecting epileptiform discharges in the EEG: real-time operation and clinical trial', In *Proc. 18th Ann. Int. Conf. of IEEE Engineering in Medicine and Biology Society, 31 Oct–3 Nov, Amsterdam.*

JUNGHÖFER, M., ELBERT, T., LEIDERER, P., BERG, P. and ROCKSTROH, B. (1997), 'Mapping EEG-potentials on the surface of the brain: A strategy for uncovering cortical sources', *Brain Topography*, Vol. 9, pp. 203–217.

KALAYCI, T. and ÖZDAMAR, Ö. (1995), 'Wavelet preprocessing for automated neural network detection of EEG spikes', *IEMB*, Vol. 14, pp. 160–166.

KALAYCI, T., ÖZDAMAR, Ö. and ERDÖL, N. (1994), 'The use of wavelet transform as a preprocessor for the neural network detection of EEG spikes', *IEEE Proceedings*, August, pp. 1–3.

KATZNELSON, R. (1981), 'EEG recording, electrode placement, and aspects of generator localization', In NUNEZ, P.L. (Ed.), *Electric fields of the brain*, Oxford University Press, New York, pp. 176–213.

KELLAWAY, P. (1979), 'An orderly approach to visual analysis: parameters of the normal EEG in adults and children', In KLAAS, D. and DALY, D. (Eds.), *Current Practice of Clinical Electroencephalography*, Raven, New York, pp. 69–147.

KENNER, H. (1976), *Geodesic Math and How to Use It*, University of California Press, Berkeley.

KIM, S., LEE, Y., KIM, J. and KIM, S. (1998), 'Automatic detection of epileptiform activity using wavelet and expert rule base', In *Proc. 20th Ann. Int. Conf. of IEEE Engineering in Medicine and Biology Society, 31 Oct–3 Nov, Hong Kong*, pp. 2078–2081.

KO, C.W. and CHUNG, H.W. (2000), 'Automatic spike detection via an artificial neural network using raw EEG data: effects of data preparation and implications in the limitations of online recognition', *Clinical Neurophysiology*, Vol. 111, pp. 477–481.

KOHONEN, T. (1989), *Self-Organization and Associative Memory, 3rd ed.*, Springer-Verlag, Berlin, Heidelberg, Germany.

KOHONEN, T. (1995), *Self-Organizing Maps*, Springer-Verlag, Berlin, Heidelberg, Germany.

KOOI, K. (1966), 'Voltage-time characteristics of spikes and other rapid electroencephalographic transients: semantic and morphological considererations', *Neurology*, Vol. 1, pp. 59–66.

LAIDLAW, J., RICHENS, A. and CHADWICK, D. (1993), *A Textbook of Epilepsy*, Longman, London.

LEHNERTZ, K. and ELGER, C. (1998), 'Can epileptiform seizures be predicted? evidence from nonlinear time series analysis of brain electrical activity', *Physical Review Letters*, Vol. 80, pp. 5019–5022.

LIPPOLD, O. (1973), *The Origin of the Alpha Rhythm*, Churchill Livingstone, Edinburgh and London.

MACGILLIVRAY, B. (1977), 'The application of automated EEG analysis to the diagnosis of epilepsy', In REMOND, A. (Ed.), *EEG Informatics. A didactic review of methods and applications of EEG data processing*, Elsevier, Amsterdam, pp. 243–261.

MAKEIG, S., BELL, A., JUNG, T.P. and SEJNOWSKI, T. (1996), 'Independent component analysis of electroencephalographic data', *Advances in Neural Information Processing Systems*, Vol. 8, pp. 145–151.

MAKEIG, S., JUNG, T.P., BELL, A. and SEJNOWSKI, T. (1997), 'Blind separation of auditory event-related brain responses into independent components', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 94, pp. 10979–10984.

MAL, S. (1999), *A Wavelet Tour of Signal Processing*, Academic Press, San Diego.

MALLAT, S. (1989a), 'Multifrequency channel decomposition of images and wavelet models', *IEEE Transactions on Signal Processing*, Vol. 37, pp. 2091–2110.

MALLAT, S. (1989b), 'A theory for multiresolution signal decomposition: the wavelet representation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, pp. 674–693.

MALLAT, S. and HWANG, W. (1992), 'Singularity detection and processing with wavelets', *IEEE Transactions on Information Theory*, Vol. 38, pp. 617–643.

MALLAT, S. and ZHANG, Z. (1993), 'Matching pursuit with frequency dictionaries', *IEEE Transactions on Signal Processing*, Vol. 41, pp. 3397–3415.

MALLAT, S. and ZHONG, S. (1992), 'Characterization of signal from their multiscale edges', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, pp. 710–732.

MALMIVUO, J., SUIHKO, V. and ESKOLA, H. (1997), 'Sensitivity distributions of EEG and MEG measurements', *IEEE Transactions on Biomedical Engineering*, Vol. 44, pp. 196–208.

MEHTA, S., ONARAL, B. and KOSER, R. (1994), 'Detection of seizure onset using wavelet analysis', In *Proc. 16th Ann. Int. Conf. of IEEE Engineering in Medicine and Biology Society, 1–6 Nov, Baltimore, Maryland.*

MOSHER, J., SPENCER, M., LEAKY, R. and LEWIS, P. (1993), 'Error bounds for EEG and MEG dipole source localization', *Electroencephalography and Clinical Neurophysiology*, Vol. 86, pp. 303–321.

NOLTE, J. (1981), *The Human Brain: An Introduction to its Functional Anatomy*, The C.V. Moshby Company, St. Louis, MO.

NUNEZ, P.L. (1981), *Electric Fields of the Brain*, Oxford University Press, New York.

NUNEZ, P.L. (1995), *Neocortical Dynamics and Human EEG Rhythms*, Oxford University Press, New York.

OGILVIE, R. (1949), *Handbook of Electroencephalography*, Addison-Wesley Press, Cambridge, Mass.

PASCUAL-MARQUI, R. (1999), 'Review of methods for solving the EEG inverse problem', *International Journal of Bioelectromegnetism*, Vol. 1, pp. 75–86.

PASCUAL-MARQUI, R., MICHEL, C. and LEHMANN, D. (1994), 'Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain', *International Journal of Psychophysiology*, Vol. 18, pp. 49–65.

PERRIN, F., BERTRAND, O. and PERNIER, J. (1987), 'Scalp current density mapping: value and estimation from potential data', *IEEE Transactions on Biomedical Engineering*, Vol. 34, pp. 283–288.

PERRIN, F., PERNIER, J., BERTRAND, O. and ECHALLIER, J. (1989), 'Spherical splines for scalp potential and current density mapping', *Electroencephalography and Clinical Neurophysiology*, Vol. 72, pp. 184–187.

PIETILÄ, T., VAPAAKOSKI, S., NOUSIAINEN, U., VÄRRI, A., FREY, H., HÄKKINEN, V. and NEUVO, Y. (1994), 'Evaluation of a computerized system for recognition of epileptic activity during long-term EEG recording', *Electroencephalography and Clinical Neurophysiology*, Vol. 90, pp. 438–443.

POPESCU, S. (1998), 'Automatic detection of interictal epileptic spikes based on digital wavelet spectrum', *European Congress on Biomedical Engineering*, Feb, pp. 2–4.

PORTER, R. (1993), 'Classification of epileptic seizures and epileptic syndromes', In LAIDLAW, J., RICHENS, A. and CHADWICK, D. (Eds.), *A Textbook of Epilepsy*, Longman, London, pp. 1–22.

QU, H. and GOTMAN, J. (1993), 'Improvement in seizure detection performance by automatic adaptation to the EEG of each patient', *Electroencephalography and Clinical Neurophysiology*, Vol. 86, pp. 79–87.

QU, H. and GOTMAN, J. (1995), 'A seizure warning system for long-term epilepsy monitoring', *Neurology*, Vol. 45, pp. 2250–2254.

QU, H. and GOTMAN, J. (1997), 'A patient-specific algorithm for the detection of seizure onset in long-term EEG monitoring: Possible use as a warning device', *IEEE Transactions on Biomedical Engineering*, Vol. 44, pp. 115–122.

QUIAN, S. and CHEN, D. (1996), *Joint Time-Frequency Analysis*, Prentice Hall, Upper Saddle River, NJ, USA.

RAMABHADRAN, B., FROST, J., GLOVER, J. and KTONAS, P. (1999), 'An automated system for epileptogenic focus localization in the eletroencephalogram', *Journal of Clinical Electroencephalography*, Vol. 16, pp. 59–68.

RAUDYS, S. and JAIN, A. (1991), 'Small sample size effects in statistical pattern recognition: Recommendations for practitioners', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, pp. 252–264.

RUPRECHT, D. and MÜLLER, H. (1998), 'A scheme for edge-based adaptive tetrahedron subdivision', In HEGE, H.C. and POLTHIER, K. (Eds.), *Visualization and Mathematics*, Springer Verlag, Heidelberg, pp. 61–70.

RUSH, S. and DRISCOLL, D. (1968), 'Current distribution in the brain from surface electrodes', *Anesthesia and Analgesia*, Vol. 47, pp. 717–723.

RUSH, S. and DRISCOLL, D. (1969), 'EEG electrode sensitivity - an application of reciprocity', *IEEE Transactions on Biomedical Engineering*, Vol. 16, pp. 15–22.

SALANT, Y., GATH, I. and HENRIKSEN, O. (1998), 'Prediction of epileptic seizures from two-channel EEG', *Medical and Biological Engineering and Computing*, Vol. 36, pp. 549–556.

SALU, Y., COHEN, L., ROSE, D., SATO, S., KUFTA, C. and HALLET, M. (1990), 'An improved method for localizing electric brain dipoles', *IEEE Transactions on Biomedical Engineering*, pp. 699–705.

SARI-SARRAF, H. and BRZAKOVIC, D. (1997), 'A shift-invariant discrete wavelet transform', *IEEE Transactions on Signal Processing*, Vol. 45, pp. 2621–2626.

SARTORETTO, F. and ERMANI, M. (1999), 'Automatic detection of epileptiform activity by single-level wavelet analysis', *Clinical Neurophysiology*, Vol. 110, pp. 239–249.

SCHIFF, S., ALDROUBI, A., UNSER, M. and SATO, S. (1994a), 'Fast wavelet transformation of EEG', *Electroencephalography and Clinical Neurophysiology*, Vol. 91, pp. 442–455.

SCHIFF, S.J., MILTON, J., HELLER, J. and WEINSTEIN, S.L. (1994b), 'Wavelet transforms and surrogate data for electroencephalographic spike and seizure localization', *Optical Engineering*, Vol. 33, July, pp. 2162–2169.

SCHIFF, S., COLELA, D., JACYNA, G., HUGHES, E., CREEKMORE, J., MARSHALL, A., BOZEK-KUZMICKI, M., BENKE, G., GAILLARD, W., CONRY, J. and WEINSTEIN, S. (2000), 'Brain chirps: spectrographic signatures of epileptic seizures', *Clinical Neurophysiology*, Vol. 111, pp. 953–958.

SEM-JACOBEN, C., BICKFORD, R., PETERSEN, M. and DODGE, H. (1953), 'Depth distribution of normal eletroencephalographic rhythms', In *Proc. Staff Meetings Mayo Clinic*, pp. 156–161.

SENHADJI, L., CARRAULT, G. and BELLANGER, J. (1994), 'Interictal EEG spike detection: A new framework based on wavelet transform', In *Proc. IEEE International Symposium on Time-Frequency and Time-Scale Analysis, 25-28 Oct, Philadelphia, PA*, August, pp. 548–551.

SENHADJI, L., DILLENSEGER, J.L., WENDLING, F., ROCHA, C. and KINIE, A. (1995), 'Wavelet analysis of EEG for three-dimensional mapping of epileptic events', *Annals of Biomedical Engineering*, Vol. 23, pp. 543–552.

SENHADJI, L., SHAMSOLLAHI, M. and BONQUIN-JEANNES, R. (1998), 'Representation of SEEG signals using time-frequency signatures', In *Proc. 20th Ann. Int. Conf. of IEEE Engineering in Medicine and Biology Society, 31 Oct–3 Nov, Hong Kong*.

SHERLOCK, B. and MONRO, D. (1998), 'On the space of orthonormal wavelets', *IEEE Transactions on Signal Processing*, Vol. 46, pp. 1716–1720.

SIDMAN, R., VINCENT, D., SMITH, D. and LEE, L. (1992), 'Experimental tests of the cortical imaging technique - applications to the response to median nerve stimulation and the localization of epileptiform discharges', *IEEE Transactions on Biomedical Engineering*, Vol. 39, pp. 437–444.

STOK, C. (1986), *The Inverse Problem in EEG and MEG with Application to Visual Evoked Reponses*, PhD thesis, University of Twente, Enschede, The Netherlands.

STOK, C. (1987), 'The influence of model parameters on EEG/MEG single dipole source estimation', *IEEE Transactions on Biomedical Engineering*, pp. 289–296.

STRANG, G. and NGUYEN, T. (1997), *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, MA 02181, USA.

SUN, M. (1997), 'An efficient algorithm for computing multishell spherical volume conductor models in EEG dipole source localization', *IEEE Transactions on Biomedical Engineering*, Vol. 44, pp. 1243–1252.

TALLON-BAUDRY, C., BERTRAND, O., DELPUECH, C. and PERNIER, J. (1996), 'Stimulus specifity of phase-locked and non-phase-locked 40 Hz visual responses in human', *The Journal of Neuroscience*, Vol. 16, pp. 4240–4249.

TARASSENKO, L., KHAN, Y. and HOLT, M. (1998), 'Identification of inter-ictal spikes in the EEG using neural network analysis', *IEE Proceedings on Scientific Measurement Technology*, Vol. 145, pp. 270–278.

THAKOR, N.V. and SHERMAN, D. (1995), 'Wavelet (time-scale) analysis in biomedical signal processing', In BRONZINO, J.D. (Ed.), *The Biomedical Engineering Handbook*, CRC Press, Boca Raton, Florida, pp. 886–906.

TOGA, A. and MAZZIOTTA, J. (2000), *Brain Mapping*, Academic Press, London.

TONG, Z., FUSHENG, Y. and QINGYU, T. (1996), 'Automatic detection and classification of epileptiform waves in EEG – a hierarchical multi-method approach', In *The 2nd IFMBE-IMIA International Workshop on Biosignal Interpretation, Japan*, IEEE Press, pp. 456–459.

TREIMAN, D. (1993), 'Status epilepticus', In LAIDLAW, J., RICHENS, A. and CHADWICK, D. (Eds.), *A Textbook of Epilepsy*, Longman, London, pp. 205–220.

TUCKER, D. (1993), 'Spatial sampling of head electrical fields: The geodesic sensor net', *Electroencephalography and Clinical Neurophysiology*, Vol. 87, pp. 154–163.

UNSER, M. (1994), 'Fast Gabor-like windowed Fourier and continuous wavelet transforms', *IEEE Signal Processing Letters*, Vol. 1, pp. 76–79.

UNSER, M. (1999), 'Splines', *IEEE Signal Processing Magazine*, Nov, pp. 22–38.

UNSER, M., ALROUBI, A. and EDEN, M. (1992), 'On the asymptotic convergence of B-spline wavelets to Gabor functions', *IEEE Transactions on Information Theory*, Vol. 38, pp. 864–972.

UNSER, M., ALDROUBI, A. and EDEN, M. (1993a), 'B-spline signal processing: Part I-theory', *IEEE Transactions on Signal Processing*, Vol. 41, pp. 821–833.

UNSER, M., ALDROUBI, A. and EDEN, M. (1993b), 'B-spline signal processing: Part II- efficient design and applications', *IEEE Transactions on Signal Processing*, Vol. 41, pp. 843–848.

UNSER, M., ALDROUBI, A. and EDEN, M. (1994a), 'Fast b-spline transforms for continuous image representation and interpolation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, pp. 277–285.

UNSER, M., ALDROUBI, A. and SCHIFF, S. (1994b), 'Fast implementation of the continuous wavelet transform with integer scales', *IEEE Transactions on Signal Processing*, Vol. 42, pp. 3519–3523.

VAPNIK, V. (1995), *The Nature of Statistical Learning Theory*, Springer-Verlag, New York.

VÄRRI, A., NEUVO, Y. and HEIKKILÄ, H. (1988), 'Computerized classification of long-term recordings of epileptic EEG', In MÄKELÄ, M., LINNAINMAA, S. and UKONEN, E. (Eds.), *Proc. of STeP. Finnish Artificial Intelligence Symposium, Helsinki, 15-18 August*, IEEE New York, NY, USA, pp. 192–197.

WALTERS, R., PRINCIPE, J. and PARK, S.H. (1989), 'Spike detection using a syntactic pattern recognition approach', *Proc. 11th Ann. Int. Conf. of IEEE Engineering in Medicine and Biology Society*, Vol. 111, pp. 1810–1811.

WANG, Y. and HE, B. (1998), 'A computer simulation study of cortical imaging from scalp potentials', *IEEE Transactions on Biomedical Engineering*, Vol. 45, pp. 724–735.

WARD, D.M. (1998), *Enhancement of Deep Epileptiform Activity in the Electroencephalogram by Adaptive Spatial Filtering*, PhD thesis, University of Canterbury, Christchurch, New Zealand.

WEBB, A. (1999), *Statistical Pattern Recognition*, Oxford University Press, New York, NY 10016, USA.

WEBBER, W., LITT, B., WILSON, K. and LESSER, R.P. (1994), 'Practical detection of epileptiform discharges (EDs) in the EEG using an artificial neural network: a comparison of raw and parameterized EEG data.', *Electroencephalography and Clinical Neurophysiology*, Vol. 91, pp. 194–204.

WILLIAMS, W., ZAVREI, H. and SACKELLARES (1995), 'Time-frequency analysis of electrophysiology signals in epilepsy', *IEEE Engineering in Medicine and Biology*, March/April, pp. 133–143.

WILSON, S., HARNER, R., DUFFY, F., THARP, B., NUWER, M. and SPERLING, M. (1996), 'Spike detection. I. correlation and reliability of human experts', *Electroencephalography and Clinical Neurophysiology*, Vol. 98, pp. 186–198.

WITKIN, A. (1983), 'Scale-space filtering', In *Proc. Int. Joint Conf. Artificial Intelligence, Karlsruhe, West-Germany*, pp. 1019–1022.

ZADEH, L. (1965), 'Fuzzy sets', *Inform. Contr.*, Vol. 8, pp. 338–353.

ZADEH, L. (1987), *Fuzzy Sets and Applications. Selected Papers by L.A. Zadeh. Edited by R.R. Yager*, John Wiley & Sons, New York.

ZADEH, L. (1992), *Fuzzy Logic for the Management of Uncertainty*, John Wiley & Sons, New York.

ZYGIEREWICZ, J., BLINOWSKA, K., DURKA, P., SZELENBERGER, W., NIEMCEWICZ, S. and ANDROSIUK, W. (1999), 'High resolution study of sleep spindles', *Electroencephalography and Clinical Neurophysiology*, Vol. 110, pp. 2136–2147.

INDEX